UNIVERSIDADE DO SUL DE SANTA CATARINA
PATRICIA RODRIGUES DE MENEZES CASTAGNA

A INTELIGÊNCIA ARTIFICIAL E O DESLOCAMENTO DO SENTIDO DE TRANSPARÊNCIA NO DISCURSO JURÍDICO

Palhoça/SC
2021

UNIVERSIDADE DO SUL DE SANTA CATARINA
PATRÍCIA RODRIGUES DE MENEZES CASTAGNA

A INTELIGÊNCIA ARTIFICIAL E O DESLOCAMENTO DO SENTIDO DE TRANSPARÊNCIA NO DISCURSO JURÍDICO

Dissertação apresentada ao Curso de Mestrado em Ciências da Linguagem da Universidade do Sul de Santa Catarina como requisito parcial à obtenção do título de Mestre em Ciências da Linguagem.

Orientadora: Profa. Dra. Solange Maria Leda Gallo

Palhoça/SC

2021

C339 Castagna, Patrícia Rodrigues de Menezes, 1976-
A inteligência artificial e o deslocamento do sentido de transparência no discurso jurídico / Patrícia Rodrigues de Menezes Castagna. – 2021.
162 f. : il. color. ; 30 cm

Dissertação (Mestrado) – Universidade do Sul de Santa Catarina, Pós-graduação em Ciências da Linguagem.
Orientação: Prof. Dra. Solange Maria Leda Gallo

1. Análise do discurso. 2. Inteligência artificial – análise do discurso. 3. Discurso jurídico. I. Gallo, Solange Leda, 1957-. II. Universidade do Sul de Santa Catarina. VI. Título.

CDD (21. ed.) 401.41

Ficha catalográfica elaborada por Alessandra Pires CRB 14/809

PATRÍCIA RODRIGUES DE MENEZES CASTAGNA

A INTELIGÊNCIA ARTIFICIAL E O DESLOCAMENTO DO SENTIDO DE TRANSPARÊNCIA NO DISCURSO JURÍDICO

Esta dissertação foi julgada adequada à obtenção do título de Mestre em Ciências da Linguagem e aprovada em sua forma final pelo Curso de Mestrado em Ciências da Linguagem da Universidade do Sul de Santa Catarina.

Palhoça, 20 de julho de 2021.

Professora e orientadora Solange Maria Leda Gallo, Doutora.
Universidade do Sul de Santa Catarina

presente por videoconferência

Professora Dra. Fernanda de Carvalho Lage, Doutora.
Universidade de Brasília

presente por videoconferência

Professora Dra. Giovanna Gertrudes Benedetto Flores, Doutora.
Universidade do Sul de Santa Catarina

Dedico esta dissertação à minha mãe, Maria Teresinha Rodrigues Menezes, e ao meu marido e parceiro de vida, Fabiano Pires Castagna, que sempre me incentivaram à busca pelo conhecimento.

AGRADECIMENTOS

Agradecer é admitir que sozinho/as nada construímos, especialmente um trabalho acadêmico interdisciplinar e complexo, a meu sentir, como o presente.

Inicio agradecendo ao Espírito Santo, que inúmeras vezes soprou em meus ouvidos os ventos da sabedoria e do conhecimento, permitindo que em meio à pandemia mundial pelo coronavírus eu permanecesse com saúde e produzisse este trabalho no tempo planejado. Sem Ele, absolutamente nada seria possível.

Agradeço à minha mãe, Maria Teresinha Rodrigues Menezes, que optou por me conceder a vida e desde criança me fez e faz acreditar em minha capacidade intelectual e na importância de seguir em frente em busca de novos conhecimentos.

Um agradecimento especial e amoroso dedico ao Fabiano Pires Castagna, marido e parceiro de vida, que sempre me incentivou ao estudo acadêmico, mesmo diante de tantas tentativas anteriores inexitosas. Nos momentos em que eu me entristecia com tanto trabalho e ausência de lazer, e ainda por entender que estava saindo muito da área jurídica, ele ressaltou a importância da formação interdisciplinar e ineditismo do tema. Obrigada por respeitar meus momentos de ausência, mesmo estando em casa, para escrever esta dissertação.

À minha sócia e amiga Vivian De Gann dos Santos agradeço por me mostrar que é possível advogar, lecionar, ser mulher e fazer um mestrado acadêmico se a agenda estiver organizada. A ela agradeço a compreensão nas ausências em nosso escritório e por me lembrar nos momentos de desânimo: “Foca no título, Paty! Segue que vai dar certo.”

À minha querida orientadora, Profa. Solange Maria Leda Gallo, sem a qual este trabalho, fruto da união entre duas áreas de conhecimento (Análise do Discurso e Direito) que até aqui não haviam se cruzado, não existiria. Com sua experiência e autoridade na teoria de base que aqui optamos (AD), me guiou durante todo o percurso, entre momentos de presença e ausência, todos profícuos e essenciais ao trabalho que ora apresentamos. Obrigada, Profa., por cada conhecimento compartilhado e tempo dedicado a este trabalho.

Às Coordenadoras do Curso de Direito da UNISUL, Profa. Virgínia Lopes Rosa e Profa. Solange Büchele de S. Thiago, bem como à Coordenadora do NPJ, Profa. Susana dos Reis Machado Pretto, pela compreensão nos momentos de adequação das cargas horárias do curso e incentivo ao mestrado acadêmico.

À coordenação do Programa de Pós-graduação em Ciências da Linguagem da UNISUL, em especial à Profa. Nádia Régia Maffi Neckel, que me acolheu com atenção e

disposição desde o primeiro dia em que entrei no PPGCL-Pedra Branca solicitando informações a respeito da seleção para o mestrado e sobre como iniciar o anteprojeto solicitado no edital.

Agradeço igualmente a todo/as o/as Professore/as do PPGCL que fizeram parte desta caminhada e pelo conhecimento que me transmitiram, permitindo que este momento tão esperado chegasse.

À UNISUL, na pessoa do Prof. Reitor Mauri Luiz Heerdt, por incentivar, mediante a concessão de bolsa parcial institucional, a participação de professores no curso de Pós-graduação da própria instituição. Apesar das dificuldades que experimentamos nos anos de 2017-2019, a bolsa concedida foi fundamental para o meu crescimento pessoal e profissional, com esperança em dias melhores. Meu sincero agradecimento.

Aos amigos que fiz durante esta agradável jornada – voltar a ser aluna e fazer novos amigos é uma satisfação – o meu abraço afetuoso, embora distante. Em 2019 nos conhecemos em sala de aula presencial (saudades), com almoços e cafés às segundas e terças, e em março de 2020 uma pandemia nos afastou, mas a amizade já estava construída, e por isso agradeço de coração aos amigo/as Nadiege Nobre Melo, Junior Laurentino e Marcelo Nicomedes, com quem aprendi muito sobre AD e os percalços da vida. Amigo/as, obrigada.

Por fim, mas não menos importante, agradeço a todo/as o/as meus/minhas aluno/as indistintamente, pois foi sem dúvida lecionando e observando o crescimento dele/as que percebi que não há tempo ideal ou idade favorável para buscarmos novos conhecimentos, mas apenas a curiosidade e o prazer em aprender. A todo/as, o meu carinho especial, sem nomeação para não cometer nenhum esquecimento.

“Você tem que pensar o seguinte, essa transparência que o procedimento padrão garante, que decorre da publicidade dos termos e tal, é um efeito de transparência, porque não existe a transparência no sentido, o sentido é sempre opaco. O que você consegue é um efeito de transparência. Esse efeito de transparência é produzido de um certo modo antes da IA, esse modo foi atropelado pela IA. Essa transparência foi ferida, conforme esse pessoal que contesta. Quem diz isso, ainda acredita na transparência, ok!” (Solange Gallo)

“É o construído do direito, né?” (Patrícia Castagna)

“Isso! É isso que você tem que vencer Patrícia acadêmica, cientista, você tem que vencer essa fé na transparência!” (Solange Gallo)

“Como assim Profa!? Vinte e um anos trabalhando com o direito e seis falando isso para os meus alunos, kkkk.” (Patrícia Castagna)

“Veja, o discurso jurídico, ele se constitui de camadas de legitimidade, que são textos matriciais, é todo um acúmulo, arquivo riquíssimo, são muitas camadas, o que acontece é que se você fere uma camada, parece que você está destruindo todo o discurso, mas é só uma camada, que depois ela vai ser reposta de outra maneira.” (Solange Gallo)

Diálogo extraído de uma das reuniões de orientação, em 25 de março de 2021, divisor de águas para a formulação deste texto.

RESUMO

Uma das bases de sustentação da teoria da Análise do Discurso é que a produção dos sentidos, em uma formação social em dado momento histórico, depende de três aspectos: a constituição, a formulação e a circulação. São três momentos inseparáveis do ponto de vista da significação, ou seja, todos os três concorrem igualmente na produção dos sentidos. (ORLANDI, 2012, p. 150-151) Sustentados nessa teoria e diante do avanço tecnológico observado no ambiente jurídico, especialmente após 2018, buscamos compreender como se constitui, como se formula e como circula o discurso de divulgação do Projeto Victor, de forma a não entrar em contradição com a hermenêutica jurídica. O Projeto Victor é o maior projeto de inteligência artificial utilizado pelo Poder Judiciário, resultante da iniciativa do Supremo Tribunal Federal em parceria com a Universidade de Brasília, iniciado no ano de 2018, tendo como uma das principais finalidades a identificação dos temas de repercussão geral de maior incidência. Assim, formulamos esta dissertação em cinco capítulos. No primeiro, introduzimos o tema e trazemos algumas noções da Análise do Discurso utilizadas ao longo do texto, bem como esclarecemos alguns aspectos tratados nos capítulos dois e três, que abordam a Hermenêutica Jurídica, a Inteligência Artificial associada ao Direito, o Projeto Victor e aplicação da IA nos Tribunais Brasileiros, e os Discursos Científico e de Divulgação Científica. No capítulo quatro, formulamos a análise por meio da definição do corpus de pesquisa e selecionamos materiais de divulgação do Projeto Victor, que começaram a ser divulgados a partir do seu nascimento, em 09/04/2018. Em seguida, recortamos vinte e quatro sequências discursivas especialmente produtoras dos efeitos que buscamos mostrar a fim de compor o corpus de análise. Com apoio do aparato teórico-metodológico da análise do discurso e na posição-sujeito analista, realizamos um gesto de leitura sobre o corpus de pesquisa e recortamos sequências extraídas do discurso de divulgação científica do Projeto Victor: em primeiro lugar, aquelas em que estão presentes uma negação, uma vez que toda negação pressupõe uma afirmação com sentido pré-construído; e, em segundo lugar, sequências que trazem a discussão sobre transparência, colocada em xeque em função da falta de publicidade dos atos que são executados pelo algoritmo do Projeto Victor. Finalizamos com o capítulo cinco, com a conclusão da análise. Quanto ao método, a análise do discurso não o estabelece como em outras áreas de conhecimento, mas preocupa-se com a finalidade: "compreender como um objeto simbólico produz sentidos" (ORLANDI, 2015, p. 64). Esta teoria permite transformar a superfície linguística em um objeto discursivo para, em seguida, iniciar-se a análise configurando-se o *corpus*, seus limites, recortes,

retomando-se conceitos e noções, de acordo com Orlandi (2015, p. 64), “pois a análise do discurso tem um procedimento que demanda um ir-e-vir constante entre teoria, consulta ao *corpus* e análise. Esse procedimento dá-se ao longo de todo o trabalho.” Pretendemos demonstrar, ao final, que não seria possível a introdução da IA nas práticas jurídicas se a verdade e a transparência, resultantes da interpretação que são produto da hermenêutica, de fato já não fossem, em alguma medida, efeitos. E é justamente nisso que está a sustentação do discurso de divulgação do projeto de IA, mas de uma forma um tanto velada, evitando expor a contradição com o discurso jurídico, no sentido de não admitir que a transparência que os interlocutores perseguem nunca existiu, e que na realidade sempre foi uma questão de legitimação política de novas práticas que procuram se inserir no direito.

Palavras-chave: Análise do Discurso. Inteligência Artificial. Discurso Jurídico.

ABSTRACT

One of the foundations of the Discourse Analysis theory is that the production of meanings, in a social formation in a given historical moment, depends on three aspects: constitution, formulation and circulation. There are three inseparable moments from the point of view of meaning, that is, all three contribute equally to the production of meanings. (ORLANDI, 2012, p. 150-151) Based on this theory and in light of the technological advance observed in the legal environment, especially after 2018, we seek to understand how the Victor Project's dissemination discourse is constituted, formulated and circulated in such a way not to contradict legal hermeneutics. The Victor Project is the largest artificial intelligence project used by the Judiciary, resulting from the initiative of the Supreme Court in partnership with the University of Brasília, started in 2018, with one of the main purposes of identifying the themes of general repercussion of higher incidence. Thus, we formulated this dissertation in five chapters. In the first, we introduce the theme and bring some notions of Discourse Analysis used throughout the text, as well as clarify some aspects dealt with in chapters two and three, which address Legal Hermeneutics, Artificial Intelligence associated with Law, Project Victor and application of AI in the Brazilian Courts, and the Scientific and Scientific Disclosure Discourses. In chapter four, we formulated the analysis by defining the research corpus and selected promotional materials for the Victor Project, which began to be disseminated from its birth, on 04/09/2018. Then, we cut out twenty-four discursive sequences that are especially producing the effects we seek to show in order to compose the corpus of analysis. With the support of the theoretical-methodological apparatus of discourse analysis and in the subject-analyst position, we performed a reading gesture on the research corpus and cut sequences extracted from the Project Victor's scientific dissemination discourse: first, those in which they are present. a negation, since every negation presupposes an affirmation with a preconstructed meaning; and, secondly, sequences that bring the discussion about transparency, put in check due to the lack of publicity of the acts performed by the Project Victor algorithm. We end with chapter five, with the conclusion of the analysis. As for the method, discourse analysis does not establish it as in other areas of knowledge, but is concerned with the purpose: “understanding how a symbolic object produces meanings” (ORLANDI, 2015, p. 64). This theory allows transforming the linguistic surface into a discursive object to then start the analysis by configuring the corpus, its limits, cuts, resuming concepts and notions, according to Orlandi (2015, p. 64) , “because discourse

analysis has a procedure that demands a constant coming and going between theory, corpus consultation and analysis. This procedure takes place throughout the entire work.” We intend to demonstrate, in the end, that it would not be possible to introduce AI in legal practices if truth and transparency, resulting from the interpretation that are the product of hermeneutics, were in fact no longer, to some extent, effects. And it is precisely in this that the AI project's disclosure discourse is supported, but in a somewhat veiled way, avoiding exposing the contradiction with the legal discourse, in the sense of not admitting that the transparency that the interlocutors pursue never existed, and which in reality has always been a matter of political legitimation of new practices that seek to be inserted in the law.

Keywords: Discourse Analysis. Artificial intelligence. Legal Discourse.

LISTA DE ABREVIAÇÕES

AD – Análise de Discurso

AAD69 – Análise Automática do Discurso 1969

AAD80 – Análise Automática do Discurso 1980

CRFB/88 – Constituição da República Federal do Brasil de 1988

IA – Inteligência artificial

FD – Formação discursiva

LGPD – Lei Geral de Proteção de Dados

OAB – Ordem dos Advogados do Brasil

MP – Ministério Público

PV – Projeto Victor

UNESCO – Organização das Nações Unidas para Educação, a Ciência e a Cultura

RG – Repercussão Geral

TSE – Tribunal Superior Eleitora

TRG - Tema de Repercussão Geral

STF – Supremo Tribunal Federal

SUMÁRIO

APRESENTAÇÃO

O ano é 2019, primeira semana de aula do semestre letivo, sala lotada de calouros do Curso de Graduação em Direito da Unisul – Campus Pedra Branca, ávidos por conhecer o que os esperava nos próximos cinco anos acadêmicos, ansiosos por me ouvir e conferir se de fato fizeram a escolha certa para o futuro profissional que os aguardava, jovens recém-saídos do ensino médio.

Repentinamente, um aluno que vinha de outro curso, mais maduro, pergunta: "Professora, diante de toda a tecnologia que nos parece também fazer parte do direito, será que daqui a cinco anos terei trabalho?" Na posição-sujeito professora que eu tomava naquele momento, impossibilitada de lhes passar insegurança, afirmei com convicção: "Claro que terá, as máquinas nunca conseguirão realizar todo o trabalho de um/a jurista." A verdade, no entanto, é que a preocupação daquele aluno, em boa parte, também era a minha. A dúvida dele era também a minha. Que habilidades o profissional do direito deve adquirir para entrar ou manter-se no mercado de trabalho? Qual será o futuro das profissões jurídicas diante do uso da inteligência artificial no trabalho dos advogados, juízes e servidores do Poder Judiciário?

Diante disso, passei a ler um pouco mais sobre o assunto tendo me deparado com o Projeto Victor, o maior projeto de IA do Poder Judiciário, resultante da iniciativa do Supremo Tribunal Federal (STF) em parceria com a Universidade de Brasília (UnB), iniciado no ano de 2018. Verifiquei que uma das finalidades do Projeto Victor era analisar o texto do processo para classificá-lo em algum tema reconhecido de repercussão geral, donde surgiu o questionamento: Como levar à Corte Superior novos temas de relevância social, política, econômica e jurídica, diferentes daqueles já existentes, se os dados textuais constantes no recurso passarão diretamente pela análise automática do Projeto Victor? Embora eu tivesse ouvido o coordenador do PV afirmar que "[...] não é o algoritmo quem decide [...]" (PEIXOTO, 2019, 22'00''a 22'20'', *podcast*), a inquietação persistia em meu pensamento, afinal os vinte anos de advocacia e cinco de docência em nível superior já tinham me revelado a complexidade do trabalho com a hermenêutica jurídica, e consequente inviabilidade de execução e/ou identificação por um algoritmo.

Foi então que em abril de 2019, por indicação do Prof. Dr. Alexandre Botelho, conheci o Programa de Pós-graduação em Ciências da Linguagem (PPGCL) da UNISUL, e desse modo conheci a Profa. Dra. Nádia Neckel, que me recebeu gentilmente, assegurou que o tema da pesquisa poderia ser desenvolvido na linha de pesquisa Texto e Discurso sob a orientação da Profa. Solange Gallo, e emprestou-me uma de suas obras particulares, "Legados de Michel

Pêcheux: inéditos em análise do discurso" (assim conheci o papa da AD), de Carlos Piovezani e Vanice Sargentini, para confecção do Anteprojeto de Dissertação. Foi um momento memorável em minha vida.

Após esse momento, conheci minha querida e sábia orientadora, Profa. Dra. Solange Gallo, na entrevista para a seleção do mestrado 2019, e após a almejada aprovação, tive a grande oportunidade de adquirir conhecimento que jamais imaginei que poderia ter acesso, culminando no estudo da Teoria da Análise do Discurso, de origem francesa, o que me possibilitaria pesquisar com maior profundidade o tema inicialmente proposto, envolvendo direito e inteligência artificial.

Defendido o projeto da dissertação em setembro de 2020, período pandêmico, mudamos o foco da pesquisa. Se antes era a forma com que os elementos são articulados aos gestos de interpretação próprios à inteligência artificial; após a qualificação do projeto, a rota foi alterada para dar espaço ao discurso de divulgação científica e, desse modo, analisarmos como se constitui, como se formula e como circula o discurso de divulgação do Projeto Victor, de forma a não entrar em contradição com a hermenêutica jurídica.

Nessa direção, desenvolvemos a pesquisa acadêmica e a análise que ora se apresenta com a finalidade de obtenção do título de Mestre em Ciências da Linguagem.

1. INTRODUÇÃO

A análise de um determinado discurso passa necessariamente pela mobilização de noções que são identificadas a partir do estudo da teoria da análise do discurso, desenvolvida por Michel Pêcheux, na França, no final da década de 1960.

Sendo o foco desta pesquisa acadêmica o discurso de divulgação científica, especialmente os efeitos de sentido de inteligência artificial produzidos pelo discurso de divulgação científica do Projeto Victor, selecionamos algumas noções da análise do discurso indispensáveis à compreensão da análise produzida no capítulo quatro.

Desse modo, iniciamos a escrita desta pesquisa, desenvolvida ao longo dos últimos dois anos, a partir da apresentação, no capítulo dois, dos princípios teóricos que sustentam a teoria da análise do discurso, que segundo Eni Orlandi (2015, p. 64), "visa compreender como um objeto simbólico produz sentidos." Por meio deles, buscamos compreender o seguinte: como se constitui, como se formula e como circula o discurso de divulgação do Projeto Victor, de forma a não entrar em contradição com a hermenêutica jurídica?

A partir daí, desdobramos a pergunta discursiva proposta e questionamos qual é o efeito de sentido de inteligência artificial produzido no discurso jurídico por meio dos textos de divulgação do Projeto Victor? Por que, no discurso jurídico, determinados sentidos são mais aceitáveis do que outros?

Selecionamos, portanto, os seguintes princípios teóricos: a) formação discursiva, memória e pré-construído; b) sujeito do discurso; c) interdiscurso e intradiscurso; e, d) condições de produção.

Considerando que o sentido não existe em si, mas é determinado pelas posições ideológicas colocadas em jogo no processo sócio-histórico em que as palavras são produzidas, a noção de formação discursiva é fundamental na análise do discurso para que se compreenda o processo de produção de sentidos, nas palavras de Eni Orlandi (2015, p. 40-41). A inscrição em uma determinada formação discursiva, e não noutra, produz um sentido e não outro, e por isso essa noção foi mobilizada na análise do discurso de divulgação científica do PV. A partir dessa noção, verificar-se-á que o sujeito divulgador interdita certos sentidos (p. ex. "o algoritmo decide") e seleciona outros (p. ex. "não é o algoritmo quem decide") a fim de que a IA seja admitida pelos atores do Direito com normalidade, como mais uma ferramenta de trabalho, que agiliza o andamento processual, e não uma tecnologia que concorra com funções públicas e impeça o acesso do jurisdicionado à justiça.

A memória e o pré-construído são noções imbricadas à formação discursiva, e desse modo, enquanto a primeira refere-se ao repetível, ao "sempre-já-lá" da interpelação ideológica, dando ao sujeito a aparência de autonomia, segundo Giovanna Flores (2014, p. 26); o segundo permite melhor perceber os entrelaçamentos entre repetição, memória e sentidos, é todo o elemento de discurso que é produzido anteriormente, em um outro discurso e independentemente, nas palavras de Freda Indursky (2011, p. 69). Assim, por exemplo, quando o sujeito divulgador do PV afirma que "não é o algoritmo quem decide...", de acordo com Peixoto (2019, 22'00''a 22'20'', *podcast*), ele retoma um enunciado de outro discurso que está pré-construído na sua fala ("o algoritmo decide").

A noção de sujeito é fundamental para a compreensão da análise de discurso, pois não existe discurso sem sujeito e, da mesma forma, não há sujeito sem ideologia, segundo Pêcheux (2014, p. 146). Será do confronto entre os sujeitos dos discursos científico, jurídico e de divulgação cientifica, afetados pela língua e pela história, que se compreenderá as contradições entre os discursos e a forma com que os sujeitos divulgadores do PV procuram consensuar a "fissura" que a inteligência artificial está produzindo no meio jurídico, pois se de um lado temos aqueles que defendem o uso da IA como um avanço para o Direito, reduzindo a tramitação de processos e concretizando o principio da eficiência administrativa, a exemplo de Toledo (2018); de outro, temos os que consideram que sem a devida transparência algorítmica, não é possível exercer controle sobre o uso da IA, o que prejudica o princípio constitucional da publicidade, nas palavras de Roque e Santos (2021, p. 69).

A distinção entre interdiscurso e intradiscurso é igualmente importante e passa pela relação entre o já-dito e o que se está dizendo, ou seja, entre a constituição do sentido e a sua formulação, segundo Orlandi (2015, p. 30). Enquanto o interdiscurso é o conjunto de formulações feitas e já esquecidas (memória) que determinam o que dizemos; o intradiscurso é a relação com que cada um de nós estabelece com essa memória (interdiscurso), pois só podemos formular se nos colocamos na perspectiva da memória discursiva. As formulações feitas pelos sujeitos divulgadores do PV (intradiscurso) sem dúvida são permeadas pelos sentidos já constituídos (interdiscurso), são noções, portanto, mobilizadas na análise realizada.

Por fim, quanto aos princípios teóricos, destacamos as condições de produção, que segundo Orlandi (2015, p. 28) compreendem fundamentalmente os sujeitos e a situação, como também a memória. A maneira como a memória "aciona" e faz valer as condições de produção, é fundamental. Um discurso, conforme Pêcheux (2014, p. 76), é sempre pronunciado a partir de condições de produção dadas, e de acordo com Orlandi (2015, p. 37), essas condições funcionam de acordo com certos fatores: relações de força, relações de sentidos e antecipação.

As condições de produção em que o discurso de divulgação científica do PV é apresentado afetam os sentidos que são produzidos em suas falas e escritas. Palestras, *podcasts*, artigos científicos, etc. são proferidos e escritos com o intuito de divulgação do PV como o nascedouro da inteligência artificial a ser absorvida como algo positivo e até mesmo revolucionário para o direito.

Esclarecida a base teórica da análise do discurso, apresenta-se a hermenêutica jurídica enquanto teoria científica da interpretação, ponto que une as duas áreas de conhecimento aqui entrelaçadas: Direito e Análise do Discurso, uma vez que qualquer indagação a respeito de hermenêutica passa, inevitavelmente, pelo estudo das relações comunicativas em sociedade, assim como pela investigação do papel desempenhado pela linguagem.

Nesse passo, além de esclarecer a hermenêutica jurídica, faz-se a distinção entre esta e a Análise de Discurso, que permite uma visão muito mais ampla, uma vez que não se reduz à fixação do sentido e alcance da norma jurídica, como o faz a hermenêutica jurídica, mas considera o próprio gesto de interpretação e objetos simbólicos que produzem sentidos, permitindo analisar, a fim de compreender, a forma com que os atores do direito articulam tal gesto quando trabalham com a inteligência artificial.

O capítulo três tem início com o panorama que atualmente encontramos envolvendo a inteligência artificial e o direito, destacando-se o uso de *machine learning* em atividades repetitivas no STF, STJ, demais Tribunais e advocacia como um todo; ferramentas de automação de documentos, pesquisa jurídica e redação de documentos; tecnologia de análise preditiva; e, a expansão das denominadas *LegalTechs* ou *Lawtechs,* termo usado para nomear *startups* que criam produtos e serviços de base tecnológica para melhorar o setor jurídico, a exemplo de soluções de *analytics* e jurimetria, desenvolvidas para ajudar advogados na tomada de decisões.

No mesmo item, expomos a normatização atualmente em vigor (Resolução n. 332/2020 do CNJ) e em discussão (Projetos de Lei n.s 5051/2019, 21/2020 e 240/2020) no Brasil, bem como o estudo e elaboração, em nível internacional, da Primeira Versão do Texto Preliminar da Recomendação sobre Ética da Inteligência Artificial (*First Draft of the Recommendation on the Ethics of Artificial Intelligence),* pela UNESCO em 2020.

Na sequência, apresentamos o Projeto Victor, o maior projeto de IA do Poder Judiciário, resultante da iniciativa do Supremo Tribunal Federal (STF) em parceria com a Universidade de Brasília (UnB), iniciado no ano de 2018, e coordenado pelo Prof. Dr. Fabiano Hartmann Peixoto. O projeto foi criado para tornar-se uma ferramenta utilizada na execução de

quatro atividades: conversão de imagens em textos no processo digital; separação do começo e do fim de um documento (peça processual, decisão, etc) em todo o acervo do STF; separação e classificação das peças processuais mais utilizadas nas atividades do STF; e, a identificação dos temas de repercussão geral de maior incidência. Apresentamos algumas etapas do seu funcionamento a as expectativas, alcançadas em boa parte de acordo com o grupo de pesquisa envolvido no projeto.

No ponto seguinte, expomos os discursos científico e de divulgação científica. Sobre este último, diz Orlandi (2012, p. 151) que é a relação estabelecida entre duas formas de discurso: o científico e jornalístico. Segundo a autora, "o jornalista lê em um discurso e diz em outro, na mesma língua." O discurso de divulgação científica é, portanto, textualização jornalística do discurso científico.

É a partir da análise do discurso de divulgação científica do Projeto Victor e demais projetos em desenvolvimento pelos Tribunais brasileiros, que se observará que há sentidos que estão sendo interditados pelos pesquisadores e sujeitos divulgadores, em especial aqueles que tratam da contradição entre o discurso científico (aplicação da IA ao direito) e o discurso jurídico; que há uma busca de sentidos que concilie o discurso jurídico, que parte da memória do Direito e produz uma verdade com raiz na norma jurídica, da interpretação, da subjetividade (BOBBIO, 2001, p. 45), e o discurso científico, que produz efeito de verdade da ciência, do algoritmo, da repetição, da objetividade (ORLANDI, 2020, p. 147).

Com apoio nessa base teórica, no capítulo quatro adentramos na análise propriamente dita, a fim de analisar os efeitos de sentido de inteligência artificial produzidos pelo discurso de divulgação científica do Projeto Victor e, para tanto, selecionamos como corpus de pesquisa materiais de divulgação do PV que começaram a ser divulgados a partir do seu nascimento, em 09/04/2018, por meio de artigos, científicos ou não, *podcasts*, bem como críticas a seu respeito, ambos não apenas no ambiente acadêmico científico, mas jurídico como um todo. Como recorte, optamos pela seleção de vinte e quatro sequências discursivas (SDs) especialmente produtoras dos efeitos que buscamos mostrar, dividindo-se em dois momentos: aquelas SDs em que estão presentes uma negação, e outras que trazem a discussão a respeito de transparência.

No que se refere ao método utilizado, esclarece-se que a análise do discurso não apresenta opções como em outras áreas de conhecimento – como a jurídica -, mas preocupa-se com a finalidade: "compreender como um objeto simbólico produz sentidos", de acordo com Orlandi (2015, p. 64). Para tanto, segundo a autora, transforma-se a superfície linguística em um objeto discursivo, em seguida, inicia-se a análise configurando-se o *corpus*, seus limites,

recortes, retomando-se conceitos e noções, “pois a análise do discurso tem um procedimento que demanda um ir-e-vir constante entre teoria, consulta ao *corpus* e análise. Esse procedimento dá-se ao longo de todo o trabalho.” É o que se pretende apresentar por meio da presente pesquisa acadêmica.

2. ANÁLISE DO DISCURSO E HERMENÊUTICA JURÍDICA

A Análise do Discurso, além de princípios teóricos próprios, conta com método específico, conforme esclarece Eni Orlandi (2015, p. 64-65):

> Nosso ponto de partida é o de que a análise do discurso visa compreender como um objeto simbólico produz sentidos. A transformação da superfície linguística em um objeto discursivo é o primeiro passo para essa compreensão. Inicia-se o trabalho de análise pela configuração do *corpus*, delineando-se seus limites, fazendo recortes, na medida mesma em que se vai incidindo um primeiro trabalho de análise, retomando-se conceitos e noções, pois a análise de discurso tem um procedimento que demanda um ir-e-vir constante entre teoria, consulta ao *corpus* e análise. Esse procedimento dá-se ao longo de todo o trabalho.

Freda Indursky (2013, p. 21), ao preparar a análise em sua obra *A fala dos quartéis e as outras vozes*, parafraseia Eni. P. Orlandi ao reafirmar que já que não se trata de uma teoria linguística que se estenda ao discurso, mas de uma teoria do discurso, já que seu objeto é específico e diferente. A AD pressupõe a linguística, mas não se limita a ela, pois sua metodologia não é adequada para tratar do objeto discursivo. O deslocamento da unidade de análise determina a necessidade de criar um corpo teórico-analítico que vise considerar a materialidade discursiva como objeto próprio (INDURSKY, 2013, p. 21).

Michel Pêcheux (2014, p. 18) afirma em *Semântica e discurso – uma crítica à afirmação do óbvio*, que o seu propósito é o de questionar as evidências fundadoras da "Semântica", tentando elaborar, na medida dos meios que dispõe, as bases de uma teoria materialista.

Desse modo, Freda Indusrsky (2013, p. 21) destaca que o conjunto de proposições teórico-analíticas inscreve-se na articulação de três regiões do conhecimento científico, determinando seu quadro epistemológico geral, segundo proposta de Pêcheux e Fuchs: a) materialismo histórico: teoria das formações sociais e suas transformações, incluindo a teoria das ideologias; b) linguística: teoria dos mecanismos sintáticos e processos de enunciação; c) teoria do discurso: teoria da determinação histórica dos processos semânticos.

Enquanto na análise do discurso investigam-se as determinações que explicam que o sentido seja aquele, mas que sempre poderia ser outro; na hermenêutica verifica-se qual é o sentido correto/verdadeiro/admitido, que não pode ser outro; e, desse modo, ao fim deste capítulo, pretende-se relacionar ambas teorias e discorrer sobre a formalidade do discurso jurídico.

2.1 ANÁLISE DO DISCURSO E PRINCÍPIOS TEÓRICOS

A análise de discurso iniciou-se na França, no final dos anos 1960, tendo como fundador Michel Pêcheux (1938-1983), então filósofo e pesquisador da *École Normale Supérieure (ENS – Paris*). Baseado em importantes estudos realizados por Canguilhem e Althusser, Pêcheux propõe a teoria da análise do discurso (BRASIL, 2011, p. 172).

A partir de Michel Pêcheux, a Ciência da Linguagem é pensada com abordagem distinta daquela proposta pelo estruturalismo, que negava o sujeito e a situação (SAUSSURE, 2006), e pela gramática gerativa transformacional (GGT), proposta por Chomsky (BRASIL, 2011, p. 172). Até então, a fala, o sujeito, o contexto, que faziam a estrutura da língua produzir sentido, eram rejeitadas, por serem consideradas acidentais e assimétricas, o que se explicava à luz de um pensamento positivista, que buscava homogeneidade, regularidade e objetividade em busca de uma ciência autônoma (FERNANDES; VINHAS, 2019, p. 134).

Entre Saussure e Pêcheux, entretanto, nas décadas de 1960 e 1970, Émile Benveniste deslocou a questão do sentido para o contexto enunciativo, sendo este autor o principal representante da teoria da enunciação. A partir da publicação dos dois tomos dos *Problemas de linguística geral* (1966 e 1974), Benveniste questionou o pensamento de Saussure no sentido de que para cada signo há apenas um significado. Para o autor, a linguagem não é somente um "eu" e um "tu" falando, mas há um contexto enunciativo que significa, sendo que a língua jamais poderá ser pensada fora de um contexto intersubjetivo.

Segundo Benveniste (1976, p. 30-31), a linguagem é um sistema simbólico especial, organizado em dois planos: a) é fato físico, pois utiliza aparelho vocal para produzir-se, do aparelho auditivo para ser percebida; e, b) é uma estrutura imaterial, uma vez que comunica significados, substituindo os acontecimentos ou as experiências pela sua "evocação". Diz ainda que a linguagem é inseparável de uma sociedade definida e particular, uma não se concebe sem a outra.

Diante de tal realidade, a análise de discurso surge a partir de questionamentos a respeito do formalismo hermético saussuriano e da negação da exterioridade, passando-se a valorizar não a frase, mas o discurso, fugindo-se da apreciação da palavra por palavra na interpretação como uma sequência fechada em si mesma (BRASIL, 2011, p. 172).

Em sua obra "O Discurso: Estrutura ou Acontecimento", Pêcheux (2012, p. 51) esclarece como a linguística é atravessada pelo discurso:

> O objeto da linguística (o próprio da língua) aparece assim atravessado por uma divisão discursiva entre dois espaços: o da manipulação de significações estabilizadas,

> normatizadas por uma higiene pedagógica do pensamento, e o de transformações do sentido, escapando a qualquer norma estabelecida a priori, de um trabalho do sentido sobre o sentido, tomados no relançar indefinido das interpretações.

Assim, "o sujeito, em detrimento do homem, é trazido para o centro da discussão. Não qualquer sujeito, mas um sujeito específico para a análise de discurso: o sujeito do inconsciente, da linguagem, interpelado pela ideologia. Um sujeito descentrado, constituído e atravessado pela linguagem." (BRASIL, 2011, p. 172)

Ao tratar da forma-sujeito do discurso, Michel Pêcheux (2014, p. 146-147) ressalta que é a ideologia que fornece as evidências pelas quais "todo mundo sabe" o que é um soldado, um operário, um patrão, uma fábrica, etc. Assim sendo:

> [...] o *sentido* de uma palavra, de uma expressão, de uma proposição etc., não existe "em si mesmo" (isto é, em sua relação transparente com a literalidade do significante), mas, ao contrário, é determinado pelas posições ideológicas que estão em jogo no processo sócio-histórico no qual as palavras, expressões e proposições são produzidas (isto é, reproduzidas). Poderíamos resumir essa tese dizendo: *as palavras, expressões, proposições etc., mudam de sentido segundo as posições sustentadas por aqueles que as empregam*, o que quer dizer que elas adquirem seu sentido em referência a essas posições, isto é, em referência às *formações ideológicas* (...). (com grifo no original).

A análise de discurso, pois, leva às últimas consequências o caráter histórico da linguagem e reestrutura o interior do próprio fazer linguístico (BRASIL, 2011, p. 172). A partir daí, Pêcheux procura explicar como as pessoas falam diferentemente (isto é, produzem diferentes sentidos) embora falem a mesma língua. O foco da AD é, portanto, o estudo dos processos históricos de produção de sentidos, o que inclui os objetos teóricos de três áreas do conhecimento: a linguística (língua), o materialismo histórico (história) e a psicanálise (sujeito) (FERNANDES; VINHAS, 2019, p. 135).

Em sua obra "Análise de Discurso: Princípios e Procedimentos", Eni Orlandi (2015, p. 23) esclarece que "a articulação dessas três regiões nos estudos do discurso é que resulta na posição crítica assumida nos anos 60 em relação à noção de leitura, de interpretação, que problematiza a relação do sujeito com o sentido (da língua com a história)."

Alguns princípios teóricos da análise do discurso, portanto, são mobilizados nesta pesquisa científica a fim de compreender qual é o efeito de sentido de "inteligência artificial" produzido no discurso jurídico por meio dos textos de divulgação do Projeto Victor? Por que, no discurso jurídico, determinados sentidos são mais "aceitáveis" do que outros?

E desse modo, apresentam-se os seguintes princípios teóricos, importantes à presente pesquisa científica: a) formação discursiva, memória e pré-construído; b) sujeito do discurso; c) interdiscurso e intradiscurso; e, d) condições de produção.

2.1.1 Formação discursiva, memória e pré-construído

Até estudar a análise de discurso, temos a firme convicção de que o que falamos são palavras nossas, fruto do nosso pensamento e convencimento. Verdadeiro equívoco. A teoria da análise de discurso nos mostra que nossos pensamentos e palavras não se constituem de forma isolada e individual, mas são resultado de muitos sentidos produzidos historicamente:

> O dizer não é propriedade particular. As palavras não são só nossas. Elas significam pela história e pela língua. O que é dito em outro lugar também significa nas "nossas" palavras. O sujeito diz, pensa que sabe o que diz, mas não tem acesso ou controle sobre o modo pelo qual os sentidos se constituem nele. Por isso é inútil, do ponto de vista discursivo, perguntar para o sujeito o que ele quis dizer quando disse "x" (ilusão da entrevista *in loco*). O que ele sabe não é suficiente para compreendermos que efeitos de sentidos estão presentificados. (ORLANDI, 2015, p. 30)

Quando nos expressamos, portanto, não há mera transmissão de informação, pois no funcionamento da linguagem, que relaciona sujeito e sentido afetado pela língua e pela história, há um complexo processo de constituição desse sujeito e produção de sentidos (ORLANDI, 2015, p. 19).

Logo, a noção de discurso, que se distancia do modo como o esquema elementar da comunicação dispõe seus elementos: emissor, receptor, código, referente e mensagem, passa pelos processos de identificação do sujeito, de argumentação, de subjetivação, de construção da realidade etc. Diz Eni P. Orlandi (2015, p. 19-20) que a linguagem serve para comunicar e não comunicar, pois as relações de linguagem são relações de sujeitos e de sentidos e seus efeitos são múltiplos e variados. O discurso, portanto, é efeito de sentido entre locutores.

É nesse sentido que ao tratar da forma-sujeito do discurso, Michel Pêcheux (2014, p. 146) esclarece que é a ideologia que fornece evidências que fazem com que uma palavra ou um enunciado "queiram dizer o que realmente dizem" e que mascaram, sob a "transparência da linguagem", aquilo que ele chama o caráter material do sentido das palavras e enunciados. No exemplo usado pelo autor, "um soldado francês não recua", é a ideologia que, por meio do "hábito" e do "uso" está designando, ao mesmo tempo, *o que é e o que deve ser:* "se você é um *verdadeiro* soldado francês, o que, de fato, você é, então você não *pode/deve* recuar".

Para explicar a afirmação acima, Pêcheux (2014, p. 146) diz que o caráter material do sentido consiste na sua dependência constitutiva no que chama de "o todo complexo das formações ideológicas", e desse modo, propõe duas teses: na primeira, desenvolve as noções de formação discursiva e processo discursivo; e, na segunda, as noções de interdiscurso e pré-construído. Neste subitem, destacam-se a formação discursiva e o pré-construído.

A primeira tese é resumida por Pêcheux (2014, p. 146-147) do seguinte modo: “as palavras, expressões, proposições etc., mudam de *sentido* segundo as posições sustentadas por aqueles que as empregam”. Ou seja, o sentido de uma palavra, de uma expressão, não “existe em si mesmo”, mas ao contrário, é determinado pelas posições ideológicas que estão em jogo no processo sócio-histórico no qual as palavras e expressões são produzidas.

A partir disso, Pêcheux (2014, p. 147) chama de *formação discursiva* (FD) aquilo que, numa formação ideológica dada, ou seja, a partir de uma posição dada numa determinada conjuntura, determinada pelo estado da luta de classes, determina o que pode e deve ser dito, o que pode ser articulado por meio de um panfleto, uma exposição, programa etc.

O sentido não existe em si, mas é determinado pelas posições ideológicas colocadas em jogo no processo sócio-histórico em que as palavras são produzidas (ORLANDI, 2015, p. 40). Por isso, a noção de formação discursiva, ainda que polêmica, é básica na AD, pois permite compreender o processo de produção de sentidos, a sua relação com a ideologia e também dá ao analista a possibilidade de estabelecer regularidades no funcionamento do discurso, ressalta Eni. P. Orlandi (2015, p. 41).

Da noção de FD, portanto, decorre a compreensão de dois pontos apresentados por Orlandi (2015, p. 41-42): a) o discurso se constitui em seus sentidos porque aquilo que o sujeito diz se inscreve em uma formação discursiva e não outra para ter um sentido e não outro, de modo que os sentidos sempre são determinados ideologicamente; b) é pela referência à formação discursiva que podemos compreender, no funcionamento discursivo, os diferentes sentidos. A palavra “terra”, por exemplo, tem sentido diverso para um indígena, um agricultor sem terra e para um grande proprietário rural. Todos os usos se dão em condições de produção – noção adiante apresentada - diferentes e referem-se a diversas formações discursivas. A autora esclarece o uso de tais ferramentas na análise realizada pelo analista de discurso:

> E isso define em grande parte o trabalho do analista: observando as condições de produção e verificando o funcionamento da memória, ele deve remeter o dizer a uma formação discursiva (e não outra) para compreender o sentido do que ali está dito. (ORLANDI, 2015, p. 41-43)

Do mesmo modo, a formação discursiva em que o sujeito se inscreve tem relação com sua memória discursiva, ponto em que Orlandi (2015, p. 27-28) observa, ao exemplificar com o enunciado da faixa negra “*Vote sem Medo!”,* encontrada em um campus universitário em época de eleições, que os sentidos não estão só nas palavras, mas na relação com a exterioridade, nas condições em que eles são produzidos e que não dependem apenas das intenções dos sujeitos. A faixa negra traz a memória do fascismo, dos conservadores, da

"direita" em sua expressão política; as palavras "sem medo" parecem apoiar o eleitor em sua posição; tomam parte contra algum dos candidatos que poderiam ameaçar os eleitores. Ou seja, a faixa negra mobiliza os sentidos do medo.

Em contrapartida, Orlandi (2015, p. 27) explica que uma faixa branca escrita em vermelho: *"vote com coragem!",* mobiliza outros efeitos de sentido. Historicamente o vermelho está ligado a posições revolucionárias, transformadoras. Sobre fundo branco, fazem apelo à vida, futuro, disposição de luta.

Se o objetivo em ambas as faixas é o mesmo, convidar o eleitor a votar, em cada uma delas observam-se diferentes filiações de sentido, remetendo-se a memórias e circunstâncias que mostram que os sentidos vão muito além das palavras, do texto, pois têm relação com a exterioridade, com as condições em que são produzidos e que não dependem só das intenções dos sujeitos. É o que Orlandi (2015, p. 29) chama de memória discursiva: o saber discursivo que torna possível todo dizer e que retorna sob a forma do pré-construído, o já-dito que está na base do dizível, sustentando cada tomada de palavra.

É nesse ponto que se observa que as noções apresentadas neste subitem se encontram imbricadas: formação discursiva, memória e pré-construído; e nesse sentido, para tratar dessa última, Giovanna G. Benedetto Flores (2014, p. 26) ressalta que a noção de formação discursiva envolve dois tipos de funcionamento: a paráfrase e o pré-construído.

O conceito de pré-construído foi introduzido na AD por Paul Henry e Michel Pêcheux em 1975, constituindo-se em ponto decisivo da teoria do discurso. Segundo Flores (2014, p. 26), o pré-construído traz os traços da memória, do repetível, do "sempre já-lá" da interpelação ideológica, dando ao sujeito enunciador a aparência de autonomia.

O próprio Pêcheux (2014, p. 89), em Semântica e Discurso, esclarece que P. Henry propôs o termo *"pré-construído"* para designar o que remete a uma construção anterior, exterior, mas sempre independente, em oposição ao que é "construído" pelo enunciado. Trata-se, portanto, do efeito discursivo ligado ao *encaixe* sintático. E mais à frente esclarece que esse efeito de pré-construído consistiria numa discrepância pela qual um elemento irrompe no enunciado como se tivesse sido pensado "antes, em outro lugar, independentemente" (PÊCHEUX, 2014, p. 142).

Segundo Freda Indursky (2011, p. 69), a noção de pré-construído permite melhor perceber os entrelaçamentos entre *repetição, memória e sentidos*. Nas palavras da autora, todo o elemento de discurso que é produzido anteriormente, em um outro discurso e independentemente, é entendido como um *pré-construído*.

O “pré-construído”, assim como as “articulações”, segundo Pêcheux (2014, p. 150-151) são elementos do interdiscurso (item 2.1.3). Enquanto o primeiro corresponde ao “sempre-já-aí” da interpelação ideológica que fornece-impõe a “realidade” e seu “sentido” sob a forma da universalidade (o “mundo das coisas”); a segunda *constitui o sujeito em sua relação com o sentido*, de modo que ela representa, no interdiscurso, aquilo que *determina a dominação da forma-sujeito*.

As noções da AD abordadas neste subitem relacionam-se à noção de sujeito do discurso, forma-sujeito e posição-sujeito, expostas a seguir.

2.1.2 Sujeito do discurso

A noção de sujeito é fundamental para a compreensão da Análise de Discurso, uma vez que, segundo essa teoria, não existe discurso sem sujeito e, da mesma forma, não há sujeito sem ideologia (PÊCHEUX, 2014, p. 146).

Ao tratar da *memória na cena do discurso,* Freda Indursky (2011, p. 68) chama a atenção para noções fundamentais que envolvem o sujeito – sujeito do discurso, repetibilidade, forma-sujeito, pré-construído, posição-sujeito, memória, formação discursiva, identificação e contra-identificação -, e observa que num dos textos fundadores da AD, que Pêcheux assina com Fuchs (2014, p. 167), a reflexão sobre sentido inicia a partir das *relações de parafrasagem* que as diferentes expressões, palavras e enunciados mantêm entre si, no interior de uma matriz de sentido, que se organiza no âmbito de uma *Formação Discursiva* (FD).

Reforça Indursky (2011, p. 70) que se a matriz de sentido se institui por meio de um *processo de repetibilidade*, coloca também limites dessa repetição, pois estabelece o que *pode e deve ser dito* – e o que não pode - no interior da FD. Ao tomar a palavra, portanto, o sujeito formula seu discurso na ilusão de que ele é a fonte de seu dizer, e assim, funciona sob o efeito do esquecimento de que os discursos pré-existem, que foram formulados em outro lugar e por outro sujeito, e que ele os retoma sem ter consciência.

A repetição do discurso, no interior de certas práticas discursivas, pode levar a um deslizamento, a uma ressignificação dos sentidos. “Isto se dá porque o sujeito do discurso pode contra-identificar-se com algum sentido regularizado ou até mesmo desidentificar-se de algum saber e identificar-se com outro.” (INDURSKY, 2011, p. 71) Mais do que isso, os sentidos, pelo trabalho que se instaura sobre a Forma-Sujeito, podem atravessar as fronteiras de determinada FD e deslizarem para outra FD, inscrevendo-se em outra matriz de sentido. É possível, portanto, a ressignificação dos sentidos, a migração dos saberes, uma vez que o

fechamento das FDs não é rígido, e sua fronteiras, porosas. Por tal razão, é possível afirmar que as FDs não existem isoladamente, mas relacionam-se entre si, construindo um complexo de formações discursivas das quais uma é dominante (PÊCHEUX, 2014, p. 149). É nesse ponto que Michel Pêcheux (2014, p. 149) propõe chamar de interdiscurso a esse "todo complexo com dominante" das formações discursivas, tratado no subitem 2.1.3 a seguir.

É interessante ressaltar nesse ponto, que o sujeito do discurso, segundo Indursky (2013, p. 42), é interpelado, mas acredita ser livre; é dotado de inconsciente, mas se percebe plenamente consciente. Uma vez constituído, o sujeito produz o seu discurso afetado pelos dois esquecimentos propostos por Pêcheux (2014, p. 161-162): o esquecimento n. 2, que se refere ao "esquecimento" pelo qual todo sujeito-falante "seleciona", no interior da formação discursiva que o domina, um enunciado, forma ou sequência, e não outro; e, o esquecimento n. 1, relacionado ao "sistema inconsciente", que dá conta do fato de que o sujeito-falante não pode, por definição, se encontrar no exterior da formação discursiva que o domina, ou seja, temos a ilusão de ser a origem do que dizemos quando, na realidade, retomamos sentidos preexistentes: é o esquecimento ideológico, de acordo com Orlandi (2015, p. 33).

No tocante à posição-sujeito na AD, Courtine (2016, p. 23 e 33) afirma que uma posição de sujeito se define como uma relação de identificação do sujeito enunciador com o sujeito universal de uma FD, que é o sujeito do saber próprio a uma FD, referindo-se ao lugar de onde se pode enunciar ("todo mundo sabe / diz / entende que...") para cada sujeito falante que venha a enunciar uma formulação a partir de um lugar inscrito em uma FD.

No mesmo texto *Definições de orientações teóricas e construção de procedimentos em Análise de Discurso*, Courtine (2016, p. 11) chama de forma-sujeito o "conjunto das diferentes posições de sujeito em uma FD como modalidades particulares da identificação do sujeito da enunciação com o sujeito do saber, com os efeitos discursivos específicos que lhes são associados."

Desse modo, a partir da constatação de Indursky (2011, p. 71) de que a memória é social, o sujeito do discurso é interpelado pelos discursos em circulação, urdidos em linguagem e tramados pelo tecido sócio-histórico que são retomados, repetidos e regularizados, e neste percurso, poderá contra-identificar-se com uma determinada FD e identificar-se com outra. É a movência dos sentidos, segundo a autora.

São esses elementos, portanto, que constituem o sujeito do discurso.

2.1.3 Interdiscurso e Intradiscurso

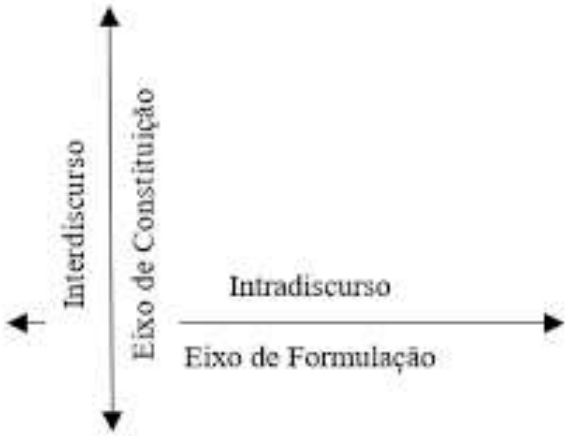
O interdiscurso, nas palavras de Orlandi (2015, p. 29), é definido como aquilo que fala antes, em outro lugar, independentemente. O interdiscurso disponibiliza dizeres que afetam o modo como o sujeito significa em uma situação discursiva dada, é o conjunto de formulações feitas e já esquecidas que determinam o que dizemos. A autora afirma que, segundo Jean-Jacques Courtine (2016), no interdiscurso fala uma voz sem nome (ORLANDI, 2015, p. 31).

No exemplo da faixa negra "Vote sem medo!", tudo o que já se disse sobre voto, eleições, eleitores e dizeres políticos que já significaram em diferentes candidatos, os sentidos da política universitária estão, de certo modo, significando ali. Experiências passadas, de ditaduras, de governos autoritários, do medo de votar, do voto que não é livre, são sentidos convocados por aquela formulação.

O interdiscurso permite verificar que o dizer não é propriedade particular, que as palavras não são só nossas, mas significam pela história e pela língua. "O sujeito diz, pensa que sabe o que diz, mas não tem acesso ou controle sobre o modo pelo qual os sentidos se constituem nele." (ORLANDI, 2015, p. 30) E o fato de que há um já-dito é fundamental para se compreender o funcionamento do discurso, a sua relação com os sujeitos e com a ideologia.

Michel Pêcheux (2014, p. 149) propõe que o próprio de toda formação discursiva é dissimular, na transparência do sentido que nela se forma, a objetividade material contraditória do interdiscurso, que reside no fato de que "algo fala" (*ça parle*) sempre "antes em outro lugar e independentemente", isto é, sob a dominação do complexo das formações ideológicas.O interdiscurso, portanto, abarca tanto as formações discursivas quanto os sentidos pré-construídos. Ao desenvolver o interdiscurso no nível das determinações discursivas, Freda Indursky (2013, p. 225) traz a noção de determinação interdiscursiva e, nessas condições, esclarece que o mesmo gesto verbal que leva o sujeito do discurso a saturar seu dizer para que este corresponda com coerência ao que pode ser dito no âmbito da FD pela qual é afetado, também o leva a definir o não dito, que permanece recalcado no interdiscurso específico de sua FD. Segundo a autora, isso demonstra que a FD não apenas indica o que pode/deve ser dito (Pêcheux), mas também o que não deve ser dito, ao que ela acrescenta também *o que pode, mas não convém ser dito nesse discurso*.

A distinção entre interdiscurso e intradiscurso passa pela relação entre o já-dito e o que se está dizendo, ou seja, entre a constituição do sentido e a sua formulação (ORLANDI, 2015, p. 30) Segundo Eni P. Orlandi, Courtine (1984) explica essa diferença representando a constituição – interdiscurso - por um eixo vertical, onde teríamos todos os dizeres já ditos e esquecidos; e, a formulação – intradiscurso – por um eixo horizontal, ou seja, aquilo que estamos dizendo naquele momento dado, em certas condições dadas:



A formulação (intradiscurso), portanto, é determinada pela relação com que cada um de nós estabelece com a nossa constituição, memória (interdiscurso), pois só podemos formular se nos colocamos na perspectiva da memória, do interdiscurso (ORLANDI, 2015, p. 31).

Pêcheux (2014, p. 154) diz que o intradiscurso, enquanto "fio do discurso" do sujeito é, a rigor, um efeito do interdiscurso sobre si mesmo, uma "interioridade" inteiramente determinada como tal "do exterior". Observa, ainda, que o interdiscurso enquanto discurso-transverso atravessa e põe em conexão entre si os elementos discursivos constituídos pelo interdiscurso enquanto pré-construído, que fornece a matéria-prima na qual o sujeito se constitui como "sujeito falante", com a formação discursiva que o assujeita.

Tudo o que falamos, portanto, resulta de algo que já foi dito e esquecido. O interdiscurso pode igualmente ser observado na área jurídica. Mesmo os posicionamentos doutrinários – fonte primeira do direito - considerados atuais, são formulados levando em conta a memória de pensamentos anteriores, tais como, a memória do Direito Romano (de 449 a.C. e 530 d.C.), da Teoria Pura do Direito (1934) de Hans Kelsen, da Teoria da Norma Jurídica (1958) de Norberto Bobbio, da Teoria Tridimensional do Direito (1968) de Miguel Reale. A criação de uma norma jurídica, desde uma emenda à constituição, até a criação de uma lei ordinária ou simples resolução, deve respeitar a pirâmide proposta por Hans Kelsen ao estabelecer que a validade da norma jurídica depende do fundamento em uma outra norma que lhe é superior.

(Fonte: http://missaodiplomatica.blogspot.com/2014/04/normas-juridicas-e-ordenamento-juridico.html)

Desse modo, desde a teoria kelseniana, respeitamos a Constituição de uma país por entender que é esta a norma hierarquicamente superior, porém, não nos damos conta, pelo esquecimento, que praticamos o positivismo jurídico criado pelas teorias clássicas do direito. Ao defender, pois, a Constituição Federal de 1988, em situações concretas na atualidade, defendemos a teoria normativa criada pela doutrina clássica positivista, este algo que fala antes, em outro lugar e tempo, de forma independente. É o interdiscurso atravessando o intradiscurso.

2.1.4 Condições de Produção

Em *Análise automática do discurso (AAD-69),* Michel Pêcheux (2014, p. 59-60) apresenta orientações conceituais para uma teoria do discurso. Inicia sua abordagem com uma crítica à ciência linguística, que se preocupa com o que fala o texto, quais as ideias principais contidas no texto, bem como se o texto está em conformidade com as normas da língua na qual se apresenta.

E já de início, ao ressaltar o deslocamento conceitual produzido por Saussure (1993), que buscou separar a homogeneidade cúmplice entre a prática e a teoria da linguagem – se a língua deve ser pensada como um *sistema,* deixa de ser compreendida como tendo a *função* de exprimir sentido – Pêcheux (2014, p. 60) afirma que "não se deve procurar o que cada parte significa, *mas quais são as regras que tornam possível* qualquer parte, quer se realize ou não." Para ele, a consequência desse deslocamento é que o "texto", de modo algum, pode ser o objeto pertinente para a ciência linguística pois ele não funciona; o que funciona é a *língua* (Pêcheux, 2014, p. 60). Com base nesta perspectiva, o autor propõe a teoria do discurso.

Pêcheux (2014, p. 76) esclarece, portanto, que um discurso é sempre pronunciado a partir de condições de produção dadas, e para tanto, extrai um exemplo a partir da análise do discurso político. Um deputado, pois, que pertence a um partido político que participa do governo ou a um partido de oposição, é porta-voz de tal ou tal grupo, que representa este ou aquele interesse, ou então está "isolado" etc. Segundo Pêcheux, bem ou mal ele está situado no interior da *relação de forças* existente entre os elementos antagonistas de um campo político dado: o que diz, o que anuncia, promete ou denuncia não tem o mesmo estatuto conforme o lugar que ele ocupa.

Por outro lado, segundo Pêcheux (2014, p. 76), o discurso deve ser remetido às *relações de sentido* nas quais é produzido: um discurso remete a outro, de modo que o processo discursivo não tem, de direito, início: "o discurso se conjuga sempre sobre um discurso prévio, ao qual ele atribui o papel de matéria-prima, e o orador sabe que quando *evoca* tal acontecimento, que já foi objeto de discurso, ressuscita no espírito dos ouvintes o discurso no qual este acontecimento era alegado (...)"

É desse modo que Pêcheux (2014, p. 78) conclui que "é *impossível analisar um discurso como um texto*, isto é, como uma sequência linguística fechada sobre si mesma, mas que é necessário referi-lo ao *conjunto de discursos possíveis* a partir de um estado definido das condições de produção (...)"

Eni P. Orlandi (2015, p. 28), nessa perspectiva, ressalta que as condições de produção compreendem fundamentalmente os sujeitos e a situação, como também a memória. A maneira como a memória "aciona", faz valer, as condições de produção, é fundamental.

Orlandi (2015, p. 37) esclarece que as condições de produção, que constituem os discursos, funcionam de acordo com certos fatores: relações de força, relações de sentidos e antecipação.

A partir da noção de relação de forças, Orlandi (2015, p. 37) destaca que o lugar a partir do qual fala o sujeito é constitutivo do que ele diz: se o sujeito fala do lugar de professor, suas palavras têm sentido diferente do que se falasse do lugar de aluno. A noção de relação de sentidos revela que não há discurso que não se relacione com outros: um discurso aponta para outros que o sustentam, assim como para dizeres futuros. Não há começo absoluto, nem ponto final para o discurso. E, por fim, o mecanismo da antecipação, segundo o qual todo o sujeito tem a capacidade de experimentar, de colocar-se no lugar em que o seu interlocutor "ouve" suas palavras, de modo a regular a argumentação, de tal forma que o sujeito dirá de um modo, ou de outro, segundo o efeito que pensa produzir em seu ouvinte.

O exercício da advocacia sustenta-se em grande parte no mecanismo de antecipação mencionado por Orlandi (2015, p. 37). Nas formas escritas ou oral, o/a advogado/a procura antecipar, colocar-se no lugar do seu interlocutor, ao preparar a sua argumentação. Isso é mais comum ainda nos julgamentos pautados nos Tribunais, em que se observa o posicionamento jurídico daquela Câmara ou Turma, faz-se contato prévio à sessão com os julgadores, para preparar a sustentação oral que ocorre minutos antes do julgamento.

Tais mecanismos de funcionamento do discurso (relações de força e de sentido e antecipação), segundo Eni P. Orlandi (2015, p. 38), repousam no que ela chama de formações imaginárias, ou seja, não são os sujeitos físicos ou os lugares como estão inscritos na sociedade que funcionam no discurso, mas suas imagens que resultam de projeções.

Esclarecidos os princípios teóricos da análise do discurso, essenciais ao desenvolvimento desta pesquisa científica, passa-se à hermenêutica jurídica e interpretação, pontos de origem da presente pesquisa.

2.2 HERMENÊUTICA JURÍDICA E INTERPRETAÇÃO

A origem da palavra hermenêutica deriva do verbo grego *hermeneuein*, usualmente traduzido por interpretar, bem como no substitutivo *hermeneia*, que designa interpretação. Aponta Ricardo Maurício Freire Soares (2012, p. 13-14) que tais vocábulos se referem à mitologia helênica, exprimindo o papel conferido ao Deus alado Hermes, o qual estava incumbido de fazer a mediação comunicativa entre os Deuses e os seres humanos. Desse modo, qualquer indagação a respeito de hermenêutica passa, inevitavelmente, pelo estudo das relações comunicativas em sociedade, assim como pela investigação do papel desempenhado pela linguagem.

Dentre as variadas expressões ou fontes do Direito, as leis positivadas, criadas pelo legislador, são formuladas em termos gerais e abstratos, a fim de que possam ser aplicadas a todos os casos da mesma espécie. Nesse sentido, esclarece André Franco Montoro (2008, p. 419):

> Passar do texto abstrato ao caso concreto, da norma jurídica ao fato real, é tarefa do aplicador do direito, seja ele juiz, tabelião, advogado, administrador ou contratante. Nessa tarefa, o primeiro trabalho consiste em fixar o verdadeiro sentido da norma jurídica e, em seguida, determinar o seu alcance ou extensão. É o trabalho de interpretação, hermenêutica ou exegese.

Dessa forma, o estudo e o conhecimento da hermenêutica é de elevadíssima importância no contexto das ciências jurídicas, “considerando ser seu objetivo específico,

exatamente, a indispensável sistematização dos processos aplicáveis, objetivando, em última análise, determinar o sentido final e o alcance específico das variadas expressões do Direito" (FRIEDE, 2015, p. 156).

Nesse ponto, destaca-se que é usual, tanto na língua portuguesa quanto em outras línguas (no alemão: *hermeneutik e auslegung*), o emprego das expressões interpretação e hermenêutica como sendo sinônimas; entretanto, não o são. Enquanto interpretar é fixar o verdadeiro sentido e alcance de uma norma jurídica (diferente da AD, como acima dito, em que se investigam as determinações que explicam que o sentido seja aquele, mas que sempre poderia ser outro); a hermenêutica, em sentido técnico, é a teoria científica da interpretação (MONTORO, 2008, p. 420).

Nas palavras de Friede (2015, p. 157), portanto:

> [...] a hermenêutica é, por via de consequência, um processo dinâmico, vivo e cíclico, que alimenta, crescente e constantemente, os próprios métodos de interpretação, procedendo, em última instância, à sistematização dos processos aplicáveis para determinar, ao final, o sentido verdadeiro e o alcance real das expressões do Direito.

Mais recentemente, ao tratar da hermenêutica e jurisprudência no Código de Processo Civil de 2015, Lei n. 13.105/2015, Oliveira (2018, p. 64-65) alerta que ainda predomina, no âmbito jurídico, a concepção de que a hermenêutica continua vinculada a um modelo clássico, que a encara como uma disciplina acessória, com função meramente auxiliar na compreensão e interpretação de textos jurídicos.

Entretanto, a hermenêutica mais contemporânea representa algo maior do que simplesmente um repositório de métodos para auxiliar o intérprete em sua tarefa de compreensão do direito. Segundo Oliveira (2018, p. 64-65), trata-se de verdadeira filosofia e não de uma disciplina acessória, mas fundante, em termos gadamerianos[1], vinculada à própria existência e sua vinculação com a linguagem.

Nesse sentido, ressalta-se o pensamento de Oliveira (2018, p. 49) a respeito da hermenêutica filosófica, que em alguns pontos se aproxima da análise de discurso, como se verá no item seguinte:

> A hermenêutica filosófica, segundo Gadamer, é algo bastante distinto daquilo que se projeta em Schleiermacher e Dilthey. Em primeiro lugar, a universalidade da hermenêutica ancora-se na linguagem e sua dimensão existencial e não em uma perspectiva formal-metodológica. Por outro lado, a subjetividade cede o lugar de protagonista para a tradição e para uma consciência que se sabe produto dos efeitos da história. Por fim, a carga pré-

[1] O autor refere-se ao filósofo Hans-Georg Gadamer (1900-2002), um dos maiores expoentes da hermenêutica. Sua obra de maior impacto foi Verdade e Método, de 1960, onde elabora uma filosofia propriamente hermenêutica, que trata da natureza do fenômeno da compreensão. Disponível em <https://pt.wikipedia.org/wiki/Hans-Georg_Gadamer > Acesso em 03 jun 2021.

> compreensiva de pré-conceitos, bem como a distância temporal que separa texto e intérprete não são obstáculos a serem superados, mas, sim, aliados deste na empreitada interpretativa.

Por outro lado, Guilhaumou (2009, p. 32) traz a perspectiva de Gadamer ao destacar que os anos 1980 foram marcados, entre os historiadores do discurso, por uma virada linguageira que se integra no que se convencionou denominar "a virada interpretativa":

> O ato de interpretar constitui, então, em uma perspectiva hermenêutica, "a arte de explicar e de transmitir aquilo que foi dito por outros e que se nos apresenta na tradição, sobretudo onde ele não é imediatamente compreensível"[2]. Trata-se, então, de situar o acontecimento linguageiro no centro da constituição linguística do mundo, lá onde se apresenta, na historicidade do discurso, uma vasta gama pragmática de atos de linguagem, nos lugares em que as línguas específicas traduzem-se umas nas outras.

Analisados os aspetos conceituais da hermenêutica jurídica enquanto teoria científica da interpretação, os três elementos que integram o conceito de interpretação são: fixação do sentido; alcance; e, norma jurídica. Para Montoro (2008, p. 420), "interpretar uma norma não é simplesmente esclarecer seus termos de forma abstrata, mas sobretudo revelar o sentido apropriado para a vida real e capaz de conduzir a uma aplicação justa". Assim, por exemplo, quando a lei estabelece a exigência de férias anuais remuneradas, busca assegurar um descanso para a saúde física e mental do trabalhador. Eis a fixação do sentido da norma jurídica.

O alcance da norma jurídica é igualmente fundamental, pois a depender dele, um dispositivo pode ser aplicado numa situação "a" ou "b". Desse modo, enquanto o Estatuto dos Servidores Públicos Federais é aplicável somente aos servidores públicos federais; a Consolidação das Leis do Trabalho estende-se ou alcança somente os empregados de empresas públicas e privadas.

A norma jurídica, por sua vez, completa o conceito de interpretação, na medida em que não são apenas as leis que necessitam de interpretação, mas igualmente os tratados, acordos ou convenções, decretos, medidas provisórias, portarias, despachos, sentenças, costumes, contratos, testamentos, etc. (MONTORO, 2008, p. 421)

Nesse ponto, observa-se que ao fixar o sentido da norma jurídica, segundo o seu alcance, o intérprete, na realidade, busca verificar o que o autor – que pode ser o legislador, o advogado, ou o juiz, de acordo com a norma jurídica objeto da interpretação – quis dizer, ou seja, analisa, interpreta o seu conteúdo; diversamente da análise de discurso que, como se verá a seguir, a partir do gesto de interpretação, analisa o que torna possível dizer "x", segundo as determinações sociais, históricas e ideológicas, e não "y".

---

[2] GADAMER, H.-G. ***L'Art de comprendre. Escrits*** II. Paris: Aubier, 1991.

Ao tratar da articulação de enunciados, implicação de propriedade e efeito de sustentação, em "Semântica e Discurso, uma crítica à afirmação do óbvio", Pêcheux (2014, p. 97-98) constata que o fenômeno da indeterminação (ou de não saturação) se encontra tanto no discurso do aparelho jurídico ("*Aquele que* causar algum prejuízo para *alguém* deve repará-lo") quanto no funcionamento cotidiano das noções gerais ("todo trabalho merece salário"), o que permite a generalidade que se espera de uma lei. E desse modo, critica o pensamento positivista de Hans Kelsen (2006):

> (...) isso significa, a nosso ver, que o jurídico não é, pura e simplesmente, um "domínio de aplicação" da Lógica, como pensam os teóricos do formalismo jurídico (Kelsen etc.), mas sim que há uma relação de simulação constitutiva entre os operadores jurídicos e os mecanismos de dedução conceptual, especialmente entre a sanção jurídica e a consequência lógica.

Desse modo, observa-se que os procedimentos e resultados adotados pela hermenêutica jurídica e análise de discurso são diversos. Como Eni P. Orlandi (2015, p. 23-24) já havia ressaltado, "o estudo do discurso distingue-se da Hermenêutica", uma vez que a Análise de Discurso questiona a própria interpretação.

Segundo Orlandi (2015, p. 24), a análise de discurso distingue-se da interpretação, pois a primeira vai muito além da segunda:

> A Análise de Discurso visa compreender como os objetos simbólicos produzem sentidos, analisando assim os próprios gestos de interpretação que ela considera como atos no domínio simbólico, pois eles intervêm no real do sentido. A Análise de Discurso não estaciona na interpretação, trabalha seus limites, mecanismos, como parte dos processos de significação. Também não procura um sentido verdadeiro através de uma "chave" de interpretação. Não há esta chave, há método, há construção de um dispositivo teórico. Não há uma verdade oculta atrás do texto. Há gestos de interpretação que o constituem e que o analista, como seu dispositivo, deve ser capaz de compreender.

Antes disso, em texto assinado por Michel Pêcheux, Jacqueline Leon, Simone Bonnafous e Jean Marie-Marandin (2014, p.251), "Apresentação da análise automática do discurso (1982)", a hermenêutica literária fora citada como uma das três ideias dominantes dos anos 1960. A problemática estruturalista, condensada em torno de Lévi-Strauss, Foucault, Barthes e Althusser era um dispositivo polêmico contra as ideias dominantes da época e nesse sentido questionava "a ideia de que o sentido dos textos é o correlato de uma consciência-leitura instalada numa subjetividade "interpretativa" sem limites."

Verifica-se, portanto, que a análise de discurso permite uma visão muito mais ampla, uma vez que não se reduz à fixação do sentido e alcance da norma jurídica, como o faz a hermenêutica jurídica, mas considera o próprio gesto de interpretação e objetos simbólicos

que produzem sentidos, permitindo analisar, a fim de compreender, a forma com que os atores do direito articulam tal gesto quando trabalham com a inteligência artificial.

3. INTELIGÊNCIA ARTIFICIAL E DIREITO

Numa das cenas do filme *Ex_Maquina*: Instinto Artificial, de 2015, o personagem de Oscar Isaac, Nathan, instiga o outro personagem, Caleb, interpretado por Domhnall Gleeson, a "ativar o intelecto" para pensar a respeito da inteligência artificial que havia criado, uma robô humanoide chamada "Ava".

Para tanto, convida-o a observar a obra de arte de Jackson Pollock, considerado o principal artista do expressionismo abstrato, que surgiu na década de 1940, o qual utilizava uma técnica chamada *dripping,* que contempla o respingo da tinta sobre a superfície, a tela. Pollock pintava com a tela no chão, baseado na ação do subconsciente. Sua arte expressava, segundo ele, os sentimentos mais escondidos e os medos, método conhecido como automatismo, que apesar do caráter espontâneo, não é aplicada de forma aleatória, pois exige detalhamento e planejamento (AGÊNCIA PAPOCA, 2020).

O diálogo entre os personagens, aos 48 minutos e 46 segundos do filme, é o seguinte:

> Nathan: E se, em vez de fazer arte sem pensar, ele decidisse não pintar nada sem saber o motivo exato? O que teria acontecido?
> Após alguns segundos pensando, responde Caleb:
> Caleb: Ele não faria nenhum traço.
> Nathan: Isso! Esse é o meu colega, que pensa antes de abrir a boca. Ele não teria feito nenhum traço. **O desafio não é agir de forma automática. É encontrar uma ação que não seja automática.** Pintar, respirar, falar, transar e se apaixonar.

O diálogo acima demonstra, no mesmo sentido, que o uso da inteligência artificial na área jurídica é também um desafio: identificar não apenas as ações que podem ser transferidas para a máquina – atualmente o maior estímulo para a pesquisa na área -, mas especialmente encontrar as ações que não sejam automáticas, que ainda necessitam da formação

jurídica e aprofundamento teórico dos atores do direito e que continuarão sendo necessárias para a produção do discurso jurídico.

Nesse ponto, é interessante ressaltar que a inteligência artificial, ao menos até o presente momento histórico, tem sido propagada como tendo a finalidade de diminuir atividades repetitivas, enfadonhas e que, por tal razão, apresentam maior possibilidade de erro se executadas pela inteligência humana (PEIXOTO, 2019). Como resultado, busca-se contribuir para a celeridade das decisões judiciais.

Esclarece Fabiano Hartmann Peixoto (2019, *podcast*) que, basicamente, se pudéssemos separar a aplicação da inteligência artificial para o Direito em dois grandes campos, separaríamos em: a) algum tipo de suporte à decisão, a *machine learning* como uma subárea da IA, que é muito apropriada como uma ferramenta de suporte à decisão, que é basicamente onde hoje se constroem as críticas e discussões principais sobre a IA e Direito; mas também há um segundo campo, mais relevante atualmente, face aos nossos problemas concretos de administração da justiça, que é b) uma atuação direta nos fluxos de processamento, a IA como uma ferramenta de apoio nos grandes problemas de processamento em termos de gestão do processo, em que se encaixa o Projeto Victor, criado pela Universidade de Brasília em parceria com o Supremo Tribunal Federal, detalhado no item 3.1.

Segundo Peixoto (2019, *podcast*), coordenador do Projeto Victor, embora o mesmo tenha componentes de decisão, não podemos chamar tal decisão de tradicionalmente jurisdicional, como os juristas trabalham no conceito do Direito, e desse modo, é possível afirmar que a IA não substituirá juízes, desembargadores, servidores da justiça. O Victor foi projetado, portanto, para atuar nos gargalos de fluxos processuais, uma aplicação clássica da IA e que permite o uso da *machine learning,* pois usa uma grande quantidade de dados e repetições e mapeamento de situações que são repetitivas, o que se observa no STF, demais Tribunais e advocacia como um todo.

Aliás, não apenas nos Tribunais a IA vem sendo pesquisada e aplicada, mas igualmente em escritórios de advocacia. “Na esfera privada, soluções baseadas em IA estão sendo incorporadas por escritórios de advocacia, que nelas enxergam potencial de otimizar tempo e reduzir custos.” (AZEVEDO, 2020) A prática jurídica – no Brasil e no mundo – está sendo transformada pela IA. Estimativas apontam que entre 23% a 35% do trabalho dos advogados podem ser automatizados, com ou sem IA.

Bernardo Azevedo (2020) afirma que as ferramentas de automação de documentos, pesquisa jurídica e redação de documentos estão dentre tecnologias nas quais o mercado

jurídico aposta, bem como existem ferramentas para aprimorar em relação à pesquisa de jurisprudência, mas quase nada sobre *eletronic discovery* (e-Discovery). Ressalta o autor que:

> A técnica de e-Discovery, que envolve pesquisar e obter informações eletrônicas em documentos, vem atraindo os olhares de empresas ao setor de tecnologia jurídica. Graças a ela, os investimentos no setor cresceram nos últimos anos. Já quanto à redação de documentos, plataformas facilitam o peticionamento (protocolo da petição depois de estar pronta), mas não a redação das peças processuais propriamente ditas.

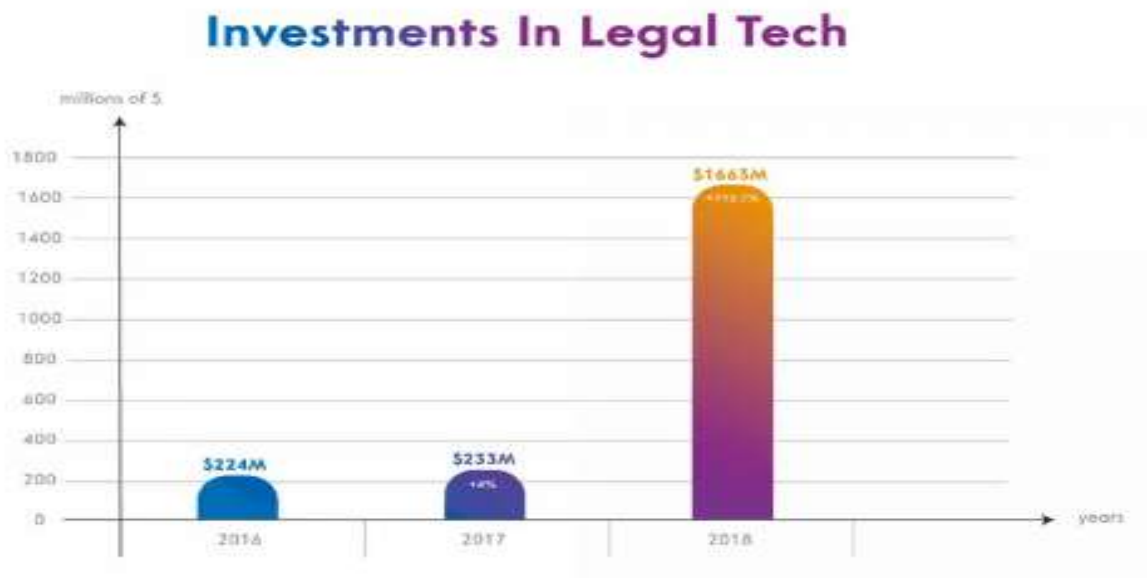
Christian A. Farmakis (2020, apud AZEVEDO) sublinha que a "menina dos olhos" dos investidores do setor de tecnologia jurídica é, certamente, a tecnologia de análise preditiva. Farmakis imagina um futuro no qual algoritmos serão capazes de identificar quão arriscado é um contrato antes mesmo de as partes o assinarem. Além disso, sistemas de inteligência artificial serão equipados para determinar se uma ação judicial terá êxito (ou não) antes mesmo de ser ajuizada.

Azevedo (2020) assevera, no entanto, que *legaltechs* brasileiras estão oferecendo soluções de *analytics* e jurimetria para ajudar advogados na tomada de decisões, mas não chegam perto do cenário proposto por Farmakis. De todo modo, é inegável que o setor de tecnologia jurídica está caminhando para tal nível de acurácia, e escritórios de advocacia equipados com *softwares* de análise preditiva têm potencial de se destacar no mercado jurídico do futuro.

Destaca Fernanda C. Lage (2021, p. 113) que na advocacia há uma expansão das denominadas *LegalTechs* ou *Lawtechs,* termo usado para nomear *startups* que criam produtos e serviços de base tecnológica para melhorar o setor jurídico. A intenção é buscar soluções para facilitar a rotina dos advogados, conectar cidadãos ao direito e aos profissionais da área jurídica.

> As Lawtechs abarcam uma série de ferramentas e processos, como: automação de elaboração de documentos jurídicos; *chatbots*; inteligência artificial preditiva; contratos inteligentes (*smart contracts*); sistemas de gerenciamento de caso; resolução de disputa eletrônica; sistemas de gestão do conhecimento e ferramentas de pesquisa inteligente. (LAGE, 2021, p. 115)

As soluções trazidas pela IA, portanto, são cada vez mais presentes na prática jurídica, reflexo inclusive do aumento de investimento em tecnologia nos últimos dez anos. Se de 2010 a 2017 o investimento total em *legaltechs* foi de US$ 1,5 bilhão, em 2018, o investimento em *startups* jurídicas superou o total de aportes realizados nos sete anos anteriores, chegando a US$ 1,6 bilhão:



Fonte: https://bernardodeazevedo.com/conteudos/como-a-inteligencia-artificial-esta-transformando-a-pratica-juridica/ apud https://news.crunchbase.com/news/legal-tech-fails-sustain-deal-counts-dollar-amounts/

Vivemos um momento, portanto, de verdadeira transformação no Direito. Entre 2018 e 2019, a prática jurídica experimentou uma verdadeira revolução no que se refere ao uso da tecnologia em atividades antes realizadas exclusivamente por advogados, juízes e servidores do Poder Judiciário. Em 2020 e 2021, com a pandemia pelo Coronavírus, esta realidade tornou-se ainda mais evidente, uma vez que o trabalho jurídico, com fóruns a portas fechadas, sobreviveu em *homeoffice* graças ao uso da tecnologia.

Do processo em papel para o processo eletrônico, antes apenas uma ferramenta, passou-se a usar a IA na redação de documentos (petições e contratos, por exemplo), e a submeter-se o recurso extraordinário, dirigido ao Supremo Tribunal Federal, a um algoritmo (Projeto Victor) que verifica se as razões recursais trazem hipótese de repercussão geral. Os prazos processuais passaram da contagem no calendário manual, ao aplicativo do celular ou ao sistema eletrônico (E-proc, PJe, Projudi, etc), que contabiliza o prazo inicial e final. A agenda, antes em papel, encontra-se no celular ou no sistema processual eletrônico vinculado à inscrição na OAB de cada advogado/a. O contato com os clientes, há poucos anos necessariamente presencial, no escritório físico do/a advogado/a, passou a ser pelas redes sociais e vídeo-chamadas, especialmente após a pandemia pelo Coronavírus. Os prazos para protocolo, antes dependentes do encerramento do expediente do serviço forense (18 ou 19 horas) e do trânsito, passaram a encerrar-se à meia-noite, quando o sistema deixa de vincular a peça protocolada ao prazo antes aberto.

Frente a toda essa realidade, a criação de normas que regulamentam as novas relações jurídicas e os conflitos a ela inerentes igualmente são motivo de discussão. A Organização das Nações Unidas para Educação, a Ciência e Cultura (UNESCO) iniciou um processo de dois anos para elaborar o primeiro instrumento mundial de definição de padrões sobre a ética da inteligência artificial (IA), após decisão tomada durante a 40ª sessão da Conferência Geral da Organização, em novembro de 2019 (UNESCO, 2020).

Tal decisão tomou como base o Estudo Preliminar sobre Ética da Inteligência Artificial (texto em anexo - A) elaborado pela Comissão Mundial da UNESCO sobre Ética do Conhecimento e Tecnologia Científica (COMEST) e apresentado em Paris em 26 de fevereiro de 2019. Esse estudo enfatiza que, atualmente, nenhum instrumento global abrange todos os campos que orientam o desenvolvimento e a aplicação da IA em uma abordagem centrada no ser humano.

Em 15 de maio de 2020, com a assistência de um Grupo de Peritos *ad hoc*, a UNESCO elaborou a Primeira Versão do Texto Preliminar da Recomendação (*First Draft of the Recommendation on the Ethics of Artificial Intelligence* - texto publicado em 07/09/2020 em anexo - B) e iniciou a fase de consultas inclusivas e multidisciplinares, que terminou em 31 de julho de 2020, para garantir que o texto preliminar seja o mais inclusivo possível.

A versão final será elaborada por um grupo de vinte e quatro especialistas de diferentes lugares do mundo, dentre os quais há um brasileiro, Edson Prestes, professor da Universidade Federal do Rio Grande do Sul (UFRGS), segundo o qual: “Esta lei é um marco para a humanidade. É a primeira em escala global sobre ética em Inteligência Artificial. Nossa expectativa é que ela venha a influenciar a elaboração de regulamentações e legislações nacionais e internacionais sobre o assunto.” (UNESCO, 2020)

Outras duas iniciativas de regulação ética da inteligência artificial foram destacadas por Humberto Martins (2020, p. 22-30), no 1º Fórum sobre Direito e Tecnologia, realizado de 29 de junho a 02 de julho de 2020, pela Fundação Getúlio Vargas (FGV). Segundo ele:

> Ambas possuem foco na ética e no fomento das tecnologias, e não na interdição de seu uso. O objetivo de ambas, portanto, é estimular as melhores práticas no campo da aplicação de inteligência artificial, em sintonia com a promoção de valores fundamentais para o desenvolvimento das sociedades. (MARTINS, 2020, p. 25)

A primeira delas, segundo Martins (2020, p. 26-27), é a “Carta Ética europeia sobre o uso de inteligência artificial em sistemas judiciais e em seu ambiente”, adotada pelo Conselho da Europa – organização internacional criada pelo Tratado de Londres em 1949, para reconstrução dos países europeus após a Segunda Guerra - em dezembro de 2018. A Carta Ética

possui cinco princípios: a) respeito aos direitos fundamentais, relacionado ao ordenamento jurídico nacional do Estado-membro e documentos internacionais que a que tenham aderido; b) não discriminação de pessoas ou grupo de pessoas; c) qualidade e segurança, por meio do uso de sistemas que sejam confiáveis, como uso de certificação; d) transparência, imparcialidade e retidão, de forma a tornar acessível e compreensível as bases pelas quais determinada decisão foi tomada com o seu uso; e, e) controle do usuário, a fim de que possam ser informados e que tenham controle sobre os potenciais e limites das escolhas às quais serão submetidos.

A segunda iniciativa refere-se à Recomendação do Conselho sobre Inteligência Artificial (OCDE/LEGAL/0449), aprovada pela OCDE (Organização para Cooperação e Desenvolvimento Econômico) em 21 de maio de 2019, que busca fomentar o uso da inteligência artificial de forma ampla, que não apenas no Poder Judiciário, como ocorre com a primeira iniciativa acima. Os princípios são similares: a) desenvolvimento inclusivo, sustentável e ao bem-estar, ou seja, a IA deve dirigir-se à qualidade de vida e inclusão das pessoas discriminadas; b) foco nos valores humanistas e equidade, de modo a proteger direitos humanos e valores democráticos como: liberdades, dignidade e autonomia dos indivíduos, proteção de dados pessoais e privacidade; c) transparência e cognoscibilidade, permitindo o esclarecimento da população sobre o que é, quais critérios de funcionamento e padrões de decisão aos potenciais afetados; d) robustez, estabilidade e segurança, que se refere à confiabilidade técnica das ferramentas e aplicações da IA; e, e) responsabilização, com foco no papel e nas ações que a IA desempenha (MARTINS, 2020, p. 29-30).

Ressalta-se que o Brasil é um dos 42 Estados, dentre membros e não membros da OCDE, que assinaram o compromisso previsto na Recomendação do Conselho sobre Inteligência Artificial (OCDE/LEGAL/0449). Portanto, a mesma servirá de base para um plano de investimentos e políticas públicas para o desenvolvimento da IA no Brasil, que está em processo de construção, de acordo com Humberto Martins (2020, p. 30).

Desse modo, de forma concreta temos atualmente no Brasil a Resolução n. 332, de 21 de agosto de 2020, do Conselho Nacional de justiça (CNJ), que dispõe sobre ética, transparência e governança na produção e no uso de Inteligência Artificial no Poder Judiciário (texto em anexo – C). Trata-se da primeira norma nacional específica em vigor sobre a matéria. De acordo com o texto, a Inteligência Artificial no âmbito do Poder Judiciário tem como principais objetivos: a promoção do bem-estar dos jurisdicionados; a realização da prestação equitativa da jurisdição; a contribuição com a agilidade e coerência do processo de tomada de decisão; a garantia da segurança jurídica; e a igualdade de tratamento aos casos absolutamente iguais. (SELEME e SOUZA, 2020)

Já aprovada e sancionada, temos a Lei Geral de Proteção de Dados (LGPD), Lei nº 13.709, de 14 de agosto de 2018, que dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural. Centrada na proteção da imagem, dados e privacidade da pessoa humana, uma das razões da criação desta lei foi o aumento do uso da inteligência artificial com finalidades ocultas voltadas aos fornecedores da relação de consumo, tais como o reconhecimento facial e uso de dados pessoais sem a autorização do consumidor.

A data de vigência desta lei já foi alterada inúmeras vezes, uma vez que envolve interesses e investimentos de empresários de todo o país. Até o fecho desta dissertação, o artigo 65, inciso I-A desta lei, incluído pela Lei n. 14.010, de 10 de junho de 2020, estabelece que os artigos 52, 53 e 54 entram em vigor no dia 1º de agosto de 2021, porém os demais artigos já se encontram em vigor.

Tem-se ainda três projetos de lei que tramitam na Câmara dos Deputados e Senado. O primeiro é o Projeto de Lei n. 5051, de 16 de setembro de 2019, de iniciativa do Senador Styvenson Valentim, do Rio Grande do Norte, que estabelece os princípios para o uso da Inteligência Artificial no Brasil.

O segundo é o Projeto de Lei n. 21, de 03 de fevereiro de 2020, de iniciativa do Deputado Federal Eduardo Bismarck, do Ceará, que cria o marco legal do desenvolvimento e uso da Inteligência Artificial (IA) pelo poder público, por empresas, entidades diversas e pessoas físicas. O texto, em tramitação na Câmara dos Deputados, estabelece princípios, direitos, deveres e instrumentos de governança para a IA, bem como o respeito aos direitos humanos e valores democráticos, tal como recomendado pelo Conselho sobre Inteligência Artificial da OCDE.

O terceiro, é o Projeto de Lei n. 240, de 11 de fevereiro de 2020, apensado ao anterior, de iniciativa do Deputado Federal Léo Moraes, de Rondônia, que propõe princípios e diretrizes para aplicação e funcionamento da Inteligência Artificial no Brasil. De acordo como art. 1º do Projeto de Lei acima, são princípios da Inteligência Artificial: I – transparência, segurança e confiabilidade; II – proteção da privacidade, dos dados pessoais e do direito autoral; III – respeito a ética, aos direitos humanos e aos valores democráticos.

O uso da IA no direito, portanto, é uma realidade e dessa forma devem as situações que a envolvem ser normatizadas a fim "de estimular as melhores práticas no campo da aplicação de inteligência artificial, em sintonia com a promoção de valores fundamentais para o desenvolvimento das sociedades" (MARTINS, 2020, p. 25). Mais do que um ponto de virada

técnico, a IA é uma disrupção tecnológica que está testando os limites da humanidade, como dito por Audrey Azoulay (2020, apud KNEBEL). Segundo a diretora geral da UNESCO, “ao entrarmos nesta nova era, devemos garantir que não sacrificaremos os nossos valores e que não falharemos em considerar as questões que essa transformação traz”.

É esta, portanto, a confluência que até junho de 2021 observamos entre a IA e o direito, uma vez que dados e informações sobre o assunto se alteram com a mesma velocidade de funcionamento do algoritmo.

Passamos, portanto, a apresentar o Projeto Victor, um dos exemplos mais recentes do uso da IA aplicada ao direito, especialmente no âmbito do Poder Judiciário e, na sequência, finalizamos este capítulo com notas sobre o discurso de divulgação científica.

3.1 O PROJETO VICTOR CRIADO PELO SUPREMO TRIBUNAL FEDERAL EM PARCERIA COM A UNIVERSIDADE DE BRASÍLIA E A APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL NOS TRIBUNAIS BRASILEIROS

Frente ao avanço da inteligência artificial no ambiente jurídico, em 2018 foi criado o Projeto Victor, o maior e mais complexo projeto de IA do Poder Judiciário e, talvez, de toda a Administração Pública, resultante da iniciativa do Supremo Tribunal Federal (STF) em parceria com a Universidade de Brasília (UnB), o que também o tornou o mais relevante Projeto Acadêmico brasileiro (STF, 2018) relacionado à aplicação de IA no Direito (STF, 2018).

De acordo com o coordenador do Projeto Victor, Fabiano Hartmann Peixoto (2020, p. 3), o início foi em 09/04/2018 com a publicação do Termo de Execução Descentralizada (TED 01/2018), expedido pelo STF. O documento foi firmado pelo Diretor-Geral do Supremo Tribunal Federal, Eduardo Silva Toledo e pela Reitora da Universidade de Brasília, Márcia Abrahão Moura.

Como resultado dessa parceria, o Supremo Tribunal Federal investiu no Projeto Victor a vultosa quantia de R$ 554.039,74, transferido à Fundação Universidade de Brasília em 09/04/2018, conforme Termo de Execução Descentralizada (TED 01/2018):

Transferências Orçamentárias (Destaques)

1º Semestre de 2018

Transferências Concedidas							
Órgão Favorecido	Fonte	NC	Data	Observação	PTRES	Grupo da Despesa	Valor - R$
CONSELHO NACIONAL DE JUSTICA	100	040001000012018NC000004	16/03/2018	DESTAQUE AO CNJ PARA , PARA PAGAMENTO DE SERVICO DE MANUTENCAO DOS VEICULOS HYUNDAI/AZERA DO STF, CONFORME TERMO DE COOPERACAOTECNICA 7/2013. OFICIOS 0510988/GSAF E 0426610/SAD/CNJ.	084441	OUTRAS DESPESAS CORRENTES	7.281,45
		040001000012018NC000007	17/05/2018	DESTAQUE AO CNJ PARA , PARA PAGAMENTO DE SERVICO DE LAVAGEM DE VEICULOS, CON- NFORME TERMO DE COOPERACAOTECNICA 7/2013. ENCAMINHAMENTO GSAF 0605838 E OFICIONº 0452970 SAD (SEI 0602697).	084441	OUTRAS DESPESAS CORRENTES	3.388,70
		040001000012018NC000008	18/05/2018	DESTAQUE AO CNJ, PARA PAGAMENTO DE SERVICO DE LAVAGEM DE VEICULOS - TERMO DE COOPERACAO TECNICA 7/2013 - PROCESSO 000374-2018.	084441	OUTRAS DESPESAS CORRENTES	1.680,00
FUNDACAO UNIVERSIDADE DE BRASILIA - FUB	100	040001000012018NC000006	09/04/2018	TRANSFERENCIA DE DOTACAO A FUNDACAO UNIVERSIDADE DE BRASILIA, CONFORME O TERMODE EXECUCAO DESCENTRALIZADA 01/2018, REFERENTE AO PROJETO DE PESQUISA E DESEN-VOLVIMENTO DE APRENDIZADO DE MAQUINA.	084435	OUTRAS DESPESAS CORRENTES	554.039,74
TRIBUNAL SUPERIOR DO TRABALHO	100	040001000012018NC000001	14/02/2018	DESTAQUE AO TST REFERENTE A DESPESAS COM COQUETEL PARA ABERTURA DE EXPOSICAO FOTOGRAFICA - ITEM 02 DA ATA 12/2017.	084435	OUTRAS DESPESAS CORRENTES	5.800,00
Total							572.189,89

Fonte: Supremo Tribunal Federal - STF, disponível em http://www.stf.jus.br/arquivo/cms/transparenciaDescentralizacaoCreditos/anexo/Transferencia 1sem2018.pdf, acesso em 12/02/2021.

Tratando-se de pesquisa científica na área de inovação, o problema específico diagnosticado pelo STF foi definido como a necessidade de uma análise de dados textuais de processos jurisdicionais para avaliar a possibilidade de uma arquitetura da inteligência artificial para classificação a ser feita em temas selecionados de repercussão geral. Em síntese, a finalidade do Projeto Victor era analisar o texto do processo para classificá-lo em algum tema reconhecido de repercussão geral. (PEIXOTO, 2020, p. 3)

Justificando-se, portanto, em princípios processuais relevantes para o direito, tais como a celeridade, eficiência e economia, bem como na proteção à saúde laboral dos serventuários do STF, que diariamente "dividiam suas respectivas jornadas entre trabalhos repetitivos, muitos dos quais enfadonhos" (PEIXOTO, 2020, p. 6), o Projeto Victor foi lançado com a finalidade de tornar-se, segundo a Ministra Carmem Lúcia (STF, 2018), uma ferramenta utilizada na execução de quatro atividades: conversão de imagens em textos no processo digital; separação do começo e do fim de um documento (peça processual, decisão, etc) em todo o acervo do Tribunal; separação e classificação das peças processuais mais utilizadas nas atividades do STF; e, a identificação dos temas de repercussão geral de maior incidência.

Para permitir o funcionamento do Projeto Victor, foi desenvolvida uma metodologia ajustada ao seu pioneirismo, que teve como fundamento o fato do desenvolvimento da pesquisa envolver conhecimentos e pesquisadores nas diferentes áreas da engenharia de software, da ciência da computação e do direito. (PEIXOTO, 2020, p. 8)

O Supremo Tribunal Federal delegou à Universidade de Brasília pelo TED 01/2018 a identificação do perfil de pesquisadores, tendo sido alocados professores, alunos de graduação

e de pós-graduação da Faculdade de Direito da UnB, da Ciência da Computação e da Faculdade de Engenharias do Gama. (PEIXOTO, 2020, p. 8-9)

Segundo Fabiano Hartmann Peixoto (2020, p. 10), inicialmente o projeto foi nomeado como Projeto de Pesquisa & Desenvolvimento de aprendizado de máquina (*machine learning)* sobre dados judiciais das repercussões gerais do Supremo Tribunal Federal – STF, porém os ministros do STF batizaram-no com o nome de Victor em homenagem à Victor Nunes Leal, Ministro do STF entre 1960 e 1969, responsável pela sistematização da jurisprudência do STF em Súmulas, prática que facilitou a aplicação de precedentes judiciais aos recursos, diminuindo a sua quantidade. Antes disso, em 07/03/2001, a Corte Suprema já havia homenageado o mesmo ministro, atribuindo-se seu nome à biblioteca da Corte: "Biblioteca Ministro Victor Nunes Leal." (LAGO, p. 367-370)

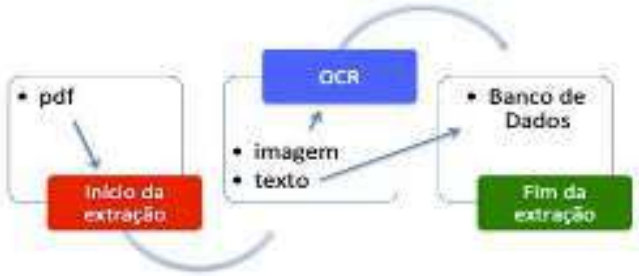
Durante o ano de 2018, as equipes de pesquisadores atuaram em dois *campi* diferentes da UnB, Gama e Darcy Ribeiro, realizando reuniões semanais para apresentação da evolução de tarefas e dificuldades encontradas. Segundo Peixoto (2020, p. 11):

> o fundamento da pesquisa foi procurar meios de melhorar as condições de trabalho (conforto e ânimo), otimizar desempenhos, assertividade de modo que tarefas repetitivas e enfadonhas sejam progressivamente apoiadas por um sistema, liberando força e tempo de trabalho para execução de atividades consideradas estratégicas pelo STF.

Diante disso, os trabalhos se voltaram para a Secretaria Judiciária do STF, ligada ao Núcleo de Repercussão Geral e à Secretaria-Geral da Presidência, a fim de agrupar dados comparativos sobre a situação do fluxo logístico processual, que indicava, ao longo de 2017, o recebimento de uma média de 400 novos processos a cada dia útil. A partir desses números, verificou-se que cerca de 1/3 da força de trabalho permanecia comprometida com o tempo despendido para uma etapa preliminar de preparação de ações de classificação, com a finalidade de classificar em temas de repercussão geral. (PEIXOTO, 2020, p. 12)

Ao mapearem a atividade humana vinculada ao objetivo principal do treinamento da máquina, os pesquisadores perceberam que algumas peças dos processos eram decisivas para a classificação dos temas de repercussão geral. Concentrados nessas peças, o produto da extração de dados combinava elementos textuais e elementos de imagem, momento em que a equipe observou que os algoritmos necessitariam de muitos ajustes, dada a complexidade do contexto de geração do pdf a partir de variados sistemas eletrônicos por todo o Brasil (PEIXOTO, 2020, p. 15). Segundo Brenno Grillo (2017), "ao todo, são mais de 40 plataformas usadas por pelos mais de 90 tribunais brasileiros, entre cortes superiores, federais, estaduais e trabalhistas."

A partir dessa etapa constituiu-se um novo banco de temas de repercussão geral (TRGs), composto por arquivos texto correspondentes ao conteúdo texto de cada arquivo pdf da base de TRGs. Após algumas mudanças e reprocessamentos de toda a base de dados até encontrar o melhor formato, os pesquisadores chegaram no seguinte modelo de extração (PEIXOTO, 2020, p. 15):



Fonte: BRASIL. Supremo Tribunal Federal. **Termo de Execução Descentralizada 01/2018**. Brasília: Supremo Tribunal Federal, 2018.

Na etapa seguinte, a equipe de pesquisadores buscou transferir para o projeto o padrão de leitura da massa de documentos jurídicos realizada pela atividade humana. Assim, criou-se uma arquitetura para classificação de peças: sentença, acórdão, recurso extraordinário (RE), agravo de recurso extraordinário (ARE) e despacho. Após esta etapa, a pesquisa foi integralmente direcionada às etapas relacionadas à classificação de temas de repercussão geral. (PEIXOTO, 2020, p. 16-17)

De acordo com Fabiano Hartmann Peixoto (2020, p. 17), em julho de 2019, após o tratamento de mais de 200 mil processos e diversas rotulagens e checagens em situações de mais de 14000 processos, chegou-se aos seguintes parâmetros de acurácia do Projeto Victor:

classe	Precisão	Recall	F1-Score
1	0,9693	0,924	0,9461
2	0,8	0,7595	0,7792
3	0,913	0,875	0,8936
4	0,8966	0,6842	0,7761
5	1	0,8462	0,9167
6	0,9589	0,9524	0,9556
7	0,8861	0,7368	0,8046
8	0,9574	0,7627	0,8491
9	0,9517	0,697	0,8047
10	0,9583	0,7931	0,8679
11	1	0,8636	0,9268
12	0,9242	0,9457	0,9349
13	0,9286	0,7222	0,8125
14	0,976	0,9606	0,9683
15	1	1	1
16	0,9559	0,942	0,9489
17	0,9658	0,8086	0,8802
18	0,9515	0,9899	0,9703
19	1	0,9767	0,9882
20	1	0,8478	0,9176
21	1	1	1
22	0,9516	0,8551	0,9008
23	1	0,9953	0,9977
24	0,9845	0,9183	0,9502
25	0,925	0,8605	0,8916
26	1	0,8037	0,8912
27	0,9881	1	0,994
28	1	0,8636	0,9268
Outras	0,9051	0,953	0,9284
Médias	**0,95681**	**0,8737**	**0,9111**

Fonte: BRASIL. Supremo Tribunal Federal. **Termo de Execução Descentralizada 01/2018**. Brasília: Supremo Tribunal Federal, 2018.

No ano de 2020 a pesquisa do PV foi ampliada, gerando-se novos desafios de curadoria de dados, classificação temática e classificação de peças a fim de incrementar a celeridade e acurácia na análise de maior quantidade de processos e situações, buscando contemplar princípios processuais como o da celeridade, eficiência e economia. (PEIXOTO, 2020, p. 19)

Em entrevista concedida em 05/10/2019, o coordenador do Projeto Victor, Fabiano Hartmann Peixoto (*podcast*, 3:30-4:30), destacou a importância da tecnologia do ponto de vista mundial, como um diferencial para as nações, posicionamento estratégico, e ressalta a oportunidade de posição internacional no âmbito das ciências sociais aplicadas e uso da inteligência artificial associada ao direito, de modo que países tradicionalmente referência em termos de pesquisa (Inglaterra, Alemanha, Estados Unidos), quando tiveram ciência do Projeto Victor, mostraram grande interesse, até porque a realidade brasileira é praticamente única no mundo em termos de volume de dados judiciais.

Frente a tal realidade, em 05/09/2019 o Ministro Dias Toffoli, então Presidente do STF, apresentou em Londres a ferramenta de Inteligência Artificial “Victor” e o Processo

Judicial Eletrônico (PJe) em palestra proferida no seminário “Novas Tendências do Direito Comum – Inteligência Artificial, Análise Econômica do Direito e Processo Civil”, que reuniu profissionais do Direito, acadêmicos do Brasil e Reino Unido.

Segundo Toffoli (2019), “o programa VICTOR, que está em fase de estágio supervisionado, promete trazer maior eficiência na análise de processos, com economia de tempo e de recursos humanos”. As tarefas que os servidores do STF levam, em média, 44 minutos, o PV fará em menos de 5 segundos. Relatou, ainda, aos ingleses que o Judiciário brasileiro sempre esteve à frente em inovação tecnológica, a exemplo da criação da urna eletrônica pela a Justiça Eleitoral, em 1996, passando pelas transmissões ao vivo das sessões do Plenário do STF, desde 2002, até a criação do “Plenário Virtual” em 2007, ambiente em que os ministros do STF julgam processos colegiadamente.

Além do Projeto Victor, desenvolvido em parceria pela UnB e STF, uma pesquisa desenvolvida pela Fundação Getúlio Vargas (FGV), por meio do Centro de Inovação, Administração e Pesquisa do Judiciário (CIAPJ/FGV), e coordenada pelo Ministro do STJ, Luis Felipe Salomão, levantou dados no sentido de que, até julho de 2020, o Poder Judiciário Brasileiro tinha ao menos 72 projetos de inteligência artificial nos tribunais, dentre eles os projetos Sócrates 1.0 e Athos, desenvolvidos no âmbito do Superior Tribunal de Justiça; Hórus, implementado no Tribunal de Justiça do Distrito Federal e Territórios; e, Secor, no Tribunal Regional Federal da 1ª Região. (SALOMÃO, 2020, p. 37)

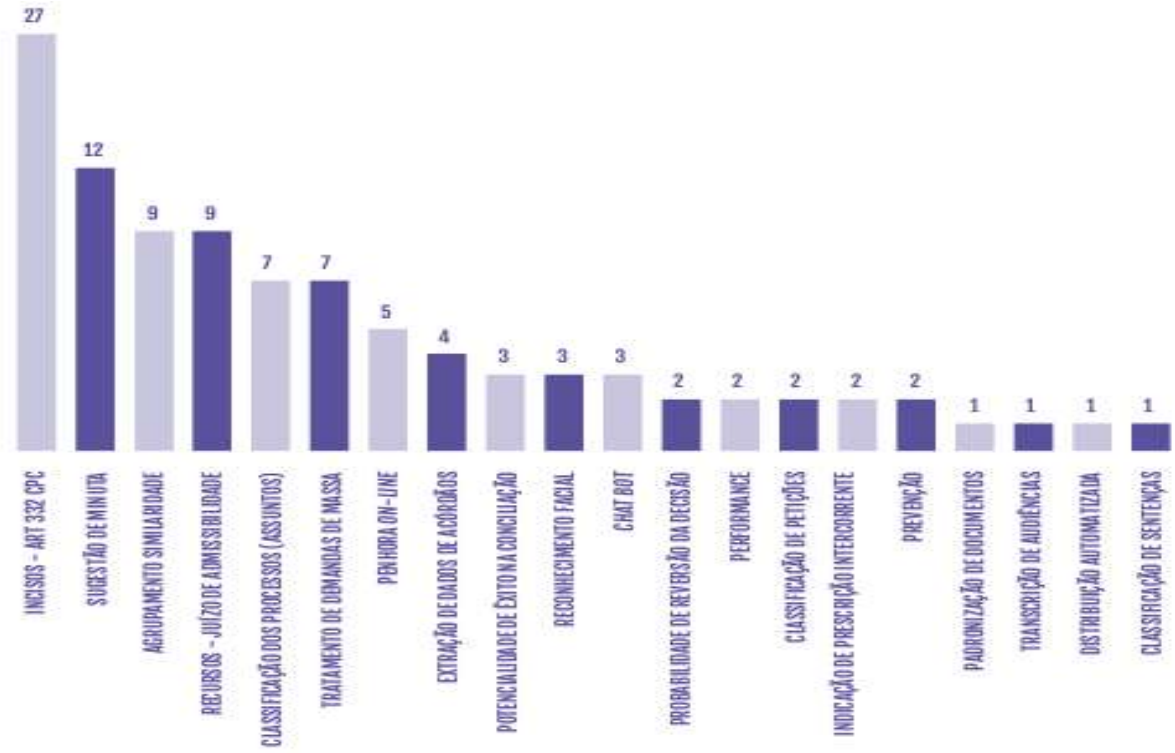
A pesquisa revelou, ainda, que 100% dos Tribunais Superiores, 100% dos Tribunais Regionais Federais (TRFs), 29% dos Tribunais Regionais do Trabalho (TRTs) e 74% dos Tribunais de Justiça (TJs) possuem sistemas de inteligência artificial já implementados ou como projetos-piloto ou em desenvolvimento (SALOMÃO, 2020, p. 38), cujos dados são demonstrados na seguinte tabela:

TRIBUNAIS	IMPLEMENTADOS	PROJETO-PILOTO	EM DESENVOLVIMENTO
Tribunais Superiores	04	03	01
TRFs	04	01	06
TRTs	01	02	04
TJs	18	04	24

Fonte: SALOMÃO, Luis Felipe (Org). **Inteligência artificial aplicada à gestão dos conflitos no âmbito do Poder Judiciário.** 1º Fórum sobre direito e tecnologia. FGV Conhecimento – Centro de

Inovação, Administração e Pesquisa do Judiciário. Publicado em 02/12/2020, p. 38. Disponível em https://ciapj.fgv.br/sites/ciapj.fgv.br/files/anais_i_forum_ia.pdf Acesso em 14 Fev 2021. (dados de julho/2020)

Observa-se, portanto, que além do pioneirismo do Projeto Victor enquanto inteligência artificial aplicada ao direito, atualmente são desenvolvidos entre o Poder Judiciário e Universidades brasileiras 72 projetos de inteligência artificial nas mais variadas atividades desenvolvidas pelos magistrados e servidores:



Fonte: SALOMÃO, Luis Felipe (Org). **Inteligência artificial aplicada à gestão dos conflitos no âmbito do Poder Judiciário.** 1º Fórum sobre direito e tecnologia. FGV Conhecimento – Centro de Inovação, Administração e Pesquisa do Judiciário. Publicado em 02/12/2020, p. 39. Disponível em https://ciapj.fgv.br/sites/ciapj.fgv.br/files/anais_i_forum_ia.pdf Acesso em 14 Fev 2021. (dados de julho/2020)

Diante da realidade apontada no que se refere à aplicação da inteligência artificial atualmente no Brasil, apresenta-se no item a seguir o discurso de divulgação científica, por meio do qual os juristas e pesquisadores produzem os sentidos para a inteligência artificial no direito, ponto fundamental para a presente pesquisa científica.

3.2 DISCURSOS CIENTÍFICO E DE DIVULGAÇÃO CIENTÍFICA

A produção dos sentidos em uma formação social em dado momento histórico depende de três aspectos, segundo Eni Orlandi (2012, p. 150-151): a constituição, a formulação

e a circulação. “São três momentos inseparáveis do ponto de vista da significação, ou seja, todos os três concorrem igualmente na produção dos sentidos. Os sentidos são como se constituem, como se formulam e como circulam.” (ORLANDI, 2012, p. 151)

Em relação ao discurso científico, não se pretende tratar, aqui, dessa formação discursiva aprioristicamente, mas apenas trazer algum conhecimento a respeito, produzido por pesquisadores que se depararam com o seu funcionamento. O primeiro que destacaremos aqui é o próprio Michel Pêcheux, que tratando dessa questão, assinala “(...) não é o Homem que produz os conhecimentos científicos, são *os homens*, em sociedade e na história, isto é *a atividade humana social e histórica*” (2014, p. 171-172), para dizer em seguida que “a história da produção dos conhecimentos não está *acima* ou *separada* da história da luta de classes.” (2014, p. 172) Ou seja, como dirá o autor, “(...) *toda* ciência é *sempre* investida (circundada e ameaçada) pelo “ideológico” (....)” (2014, p. 183)

Por essa razão, a verdade no discurso científico é também um efeito, pois está associada a fatores históricos, sociais e ideológicos. Uma vez que o cientista não produz nada sozinho, mas sim em sociedade, em grupo de pesquisadores, a sua atividade é social e abrangente, resultante de um contexto social e historicamente construído, e não particular. Se a ciência se desenvolve e uma determinada direção e não noutra, não é livre de ideologia. Ou seja, o que se afirma na ciência é ideologicamente marcado como em qualquer outro discurso, como propõe Valéria T. S. Adinolfi (2007, p. 7):

> Aqui voltamos ao discurso científico, constituído como uma metalinguagem que silencia os demais discursos possíveis. Na ilusão de saberes cristalizados, a-históricos, universais, neutros e objetivos a ciência se constitui, estabelecendo uma linguagem que pretensamente traz as mesmas características. A comunidade científica é o lugar do estabelecimento desses sentidos, e se constitui uma formação científica com um regime de produção de verdade científica à qual o cientista se assujeita. É pela assimilação de técnicas e procedimentos válidos para a obtenção e produção da verdade, pelo treinamento no uso e reprodução da metalinguagem científica, que se constitui enquanto cientista.

Ao tratar da construção do discurso científico, Dayse C. Bersot e Jacqueline C. P. Lima (2012, p. 293) ressaltam que:

> Ao se constituir, o discurso científico apaga as marcas dos outros discursos possíveis e da historicidade na formação dos sentidos, de onde vem a ilusão de universalidade. Ao fazê-lo, a história é silenciada e ressurge como um discurso pronto, acabado, a-histórico, mediando à relação do cientista com o mundo através da linguagem, determinando os sentidos de sua fala, filiando-o a uma formação discursiva própria, caracterizando-o, interpelando-o enquanto sujeito subordinado às regras dessa formação discursiva.

É nesse sentido que diremos, aqui, que assim como no discurso jurídico, também no discurso científico, a verdade e a transparência são efeitos de sentido, resultantes de um longo processo histórico de legitimação de um certo dizer, em detrimento de outros.

A análise de discurso leva em conta, portanto, a história: não há sentidos já dados, estes são constituídos por sujeitos inscritos na história num processo simbólico, duplamente descentrado pelo inconsciente e pela ideologia. “Os sujeitos têm um papel ativo, determinante na constituição dos sentidos, mas este processo escapa ao seu controle consciente e às suas intenções.” (ORLANDI, 2020, p. 140)

Desse modo, a produção de determinados sentidos - o que também ocorre no discurso de divulgação científica - é pautada pelas condições de produção de cada sujeito que formula e, segundo Eni Orlandi (2020, p. 138), tem relação direta com a interpretação. Afirma a autora:

> A relação com o simbólico, como tenho proposto, é uma relação com a intepretação. Ela está na base da própria constituição de sentido, já que, diante de qualquer objeto simbólico, o sujeito é instado a interpretar (a dar sentido) determinado pela história, pela natureza do fato simbólico, pela língua. Aí está o princípio mesmo da ideologia: não há sentido sem interpretação mas este processo de constituição de sentido (sua historicidade) não é transparente para o sujeito. Ao contrário, é através de um processo imaginário que o sentido se produz no sujeito na relação que interliga linguagem/pensamento/mundo. (ORLANDI, 2020, p. 138)

O discurso científico, no entanto, é diverso do discurso de divulgação científica. De acordo com Dominique Maingueneau (1997, p. 57), no discurso científico, “a tendência é fazer coincidir o público de seus produtores com o de seus consumidores: escreve-se apenas para seus pares que pertencem a comunidades restritas e de funcionamento rigoroso”, o que aliás, também ocorre no discurso jurídico. Já o discurso de divulgação científica, segundo José Horta Nunes (2003, p. 44), implica o direcionamento para um público que não coincide com o dos cientistas, ou seja, “seria uma prática de reformulação de um discurso-fonte em um discurso segundo, que compreende tradução, resumo, resenha, textos pedagógicos direcionados a tal ou tal grupo social.”

Entretanto, Eni Orlandi (2012, p. 151) afirma que não se trata de tradução, pois a divulgação científica é a relação estabelecida entre duas formas de discurso – o científico e jornalístico – na mesma língua e não entre duas línguas. Segundo a autora, “o jornalista lê em um discurso e diz em outro, na mesma língua.” O discurso de divulgação científica é, portanto, textualização jornalística do discurso científico.

Segundo Solange Gallo (2011, p. 667), é a partir da formação discursiva do sujeito do discurso que fala “sobre”, que se determina o dizer jornalístico, de modo que a ciência (ou

qualquer outro tema tratado) se converte em notícia, algo que precisamos saber enquanto cidadãos – é, por exemplo, o que lemos a respeito de tratamentos e vacinas contra a Covid-19. Esclarece a autora que "não se trata, portanto, de uma troca de conhecimento entre sujeitos inscritos em um mesmo discurso, mas sim de um dizer unilateral da parte do jornalista, que cita o dizer do cientista, corroborando com isso seu próprio dizer." (GALLO, 2011, p. 667)

Todo esse processo, que passa do discurso científico para o discurso de divulgação científica, tem como resultado algo maior que interfere na sociedade, que Orlandi (2012, p. 152) chama de efeito de "exterioridade" da ciência. Em suas palavras, "a ciência sai de si, sai de seu próprio meio para ocupar um lugar social e histórico no cotidiano dos sujeitos, ou seja, ela vai ser vista como afetando as coisas a saber no cotidiano da vida social."

De todo modo, nas diferentes ordens do discurso – científico, religioso, jurídico, etc. – há diferentes modos de interpretação. "Ainda quando há interdição da interpretação, há espaço de trabalho do sujeito e da história na relação com os sentidos." (ORLANDI, 2020, p. 143) É o que ocorre, por exemplo, no discurso da matemática, cujo efeito é o de representar a não-relação com a exterioridade. Apesar disso, ao especificar o que não se pode dizer, já se produz um gesto de interpretação mínimo.

É o que se observa igualmente nos efeitos de sentido de inteligência artificial que vem sendo produzidos pelo discurso de divulgação científica do Projeto Victor (PV) e demais projetos em desenvolvimento pelos Tribunais brasileiros. Há sentidos que estão sendo interditados pelos pesquisadores, em especial aqueles que tratam da contradição entr2e o discurso científico (aplicação da IA ao direito) e o discurso jurídico. Conforme textos de artigos científicos e *podcast* a serem analisados no capítulo seguinte, o que se observa é uma busca de sentidos, pelos pesquisadores do PV, que concilie o discurso jurídico, que parte da memória do Direito e produz uma verdade com raiz na norma jurídica, da interpretação, da subjetividade (BOBBIO, 2001, p. 45), e o discurso científico, que produz efeito de verdade da ciência, do algoritmo, da repetição, da objetividade (ORLANDI, 2020, p. 147).

O discurso de divulgação científica, produzido para divulgar o Projeto Victor, aproxima-se do discurso das novas tecnologias digitais, que se constitui, segundo Dias e Couto (2011, p. 644), pela filiação aos sentidos de inovação, avanço tecnológico, novidade, inclusão, internet, redes sociais e outros. Aproxima-se também do discurso publicitário, que procura omitir as equivocidades do discurso científico, apresentando somente o que é favorável. Tais efeitos de sentido, vistos em geral como inovadores, positivos e bem-vindos em todas as áreas de conhecimento, permitem e facilitam, para os pesquisadores do PV, produzir efeitos de sentido conciliáveis com o direito, que contrariamente, tem como característica principal, em

especial no âmbito dos Tribunais, a morosidade e a ineficiência dos processos ajuizados a fim de solucionar conflitos, como será exposto a seguir.

Desse modo, como explica Orlandi (2012, p. 157), o discurso de divulgação científica pode ser visto como uma *versão* do texto científico, ele parte de um texto que é da ordem do discurso científico e, pela textualização jornalística, organiza sentidos de modo a manter um efeito-ciência. Ocorre aí o que a autora chama de "didatização do discurso da ciência." (ORLANDI, 2012, p. 158)

É o que se observa com o Projeto Victor. A Universidade de Brasília, por meio de seus pesquisadores das áreas de Direito e Tecnologia – das Faculdades de Direito, Ciência de Computação e Engenharias do Gama – produzem o texto científico a respeito do projeto (Brasil, STF, Termo de execução Descentralizada 01/2018, 2018); enquanto o coordenador do PV e demais pesquisadores envolvidos divulgam-no por meio da publicação de livros, artigos – científicos ou não -, seminários e palestras por todo o país, entrevistas por *podcasts*, tendo como finalidade produzir os efeitos de sentido para a inteligência artificial no direito, algo relativamente novo, que está sendo desenvolvido desde 2018, com o nascimento deste projeto.

Nesse ponto é importante distinguir a posição sujeito-jornalista da posição sujeito que produz informação. Enquanto o primeiro produz informação relacionada à notícia jornalística (discurso jornalístico), o segundo produz informação não necessariamente enquanto notícia (discurso de divulgação científica). É o que faz o sujeito-coordenador do Projeto Victor, que produz informação sobre o projeto, mas não na forma de notícia. A produção de informação, portanto, não necessariamente é feita por um sujeito-jornalista, quando está formulada no discurso de divulgação.

No tocante à informação, Solange Gallo (2018, p. 349) esclarece em capítulo de livro intitulado Discurso e Novas Tecnologias de Informação que:

> Tenho pensado a informação como um dizer que se produz em uma determinada discursividade, e que dela é retirado para ser transportado para outra discursividade, perdendo, nesse movimento, sentidos pré-construídos. Uma vez inserido na nova discursividade, outros sentidos pré-construídos serão mobilizados para a interpretação do enunciado transposto, que aí é interpretado, por essa razão, como "informação". Assim, por exemplo, um enunciado que é produzido no discurso científico, ao ser inserido no discurso jornalístico, produz aí o sentido de "informação científica". O que permite pensar que os sujeitos que compartilham saberes de uma mesma formação discursiva, produzem o conhecimento que é coletivo e válido para todo sujeito que aí se inscreve. Por outro lado, quando o sujeito não compartilha esses saberes, a interpretação de certos enunciados sofre um deslocamento, transformando-se em outro, como é o caso do enunciado científico que se transforma em "informação científica" (notícia) para o sujeito inscrito no discurso jornalístico.

Segundo propõe Gallo (2018, p. 349), a informação é uma porção de um texto que vem de outro discurso (PV vem do discurso científico, por exemplo) e é colocado no discurso de divulgação científica. Porém, nesse movimento, o texto original perde os pré-construídos que havia no discurso científico – quando lá se explica, por exemplo, o funcionamento do algoritmo no PV que interpreta determinados textos jurídicos – de modo que o sujeito que produz a informação (sujeito-coordenador), traz somente o trecho que pretende divulgar para então dizer qual o sentido com as suas palavras, e para tanto utiliza outros pré-construídos, aqueles mais próximos do senso comum, para que um sujeito que não seja da área de tecnologia possa compreender, já que seus interlocutores não são cientistas, mas juristas.

Essa passagem do texto, portanto, de um discurso para o outro (científico para divulgação científica) faz com que o sujeito-coordenador produza informação sobre o Projeto Victor e ao produzi-la ele está divulgando não apenas o projeto, mas a viabilidade do uso da inteligência artificial no campo jurídico para os profissionais do direito.

Os textos, portanto, produzidos pelo sujeito-coordenador e demais pesquisadores do PV, que são utilizados como material de divulgação do projeto, constituem o corpus de pesquisa do presente trabalho acadêmico, pois são eles que produzem um certo sentido de inteligência artificial, quando tantos outros seriam possíveis a partir do discurso de origem: o discurso científico que produz inteligência artificial através da proposição de algoritmos e cálculos.

O recorte dessa análise prevê, ainda, uma aproximação dos sentidos produzidos em respeito à inteligência artificial, a sentidos próprios de uma interpretação de natureza hermenêutica, procurando identificar encadeamentos, rupturas, etc. É o que será trabalhado na análise, objeto do próximo capítulo.

Dito em outras palavras, é no discurso de divulgação científica que se encontra a discussão, em um âmbito social mais amplo, do projeto de introdução da inteligência artificial no Poder Judiciário brasileiro, e de todas as consequências que isso pode acarretar. Por esse motivo, foi nesse discurso que buscamos constituir nosso corpus, pois é nessa formação discursiva que estão os sentidos que queremos analisar, referentes às contradições inerentes à essa empreitada. Este será nosso recorte neste corpus.

4. ANÁLISE: OS EFEITOS DE SENTIDO DE INTELIGÊNCIA ARTIFICIAL PRODUZIDOS PELO DISCURSO DE DIVULGAÇÃO CIENTÍFICA DO PROJETO VICTOR

A análise de discurso desenvolve-se especialmente a partir de um gesto de interpretação delimitado em recortes. A definição do corpus já é um gesto de interpretação, do mesmo modo que os recortes que se fazem no percurso da análise. A seleção do material de divulgação do Projeto Victor como corpus desta pesquisa é, portanto, um gesto de interpretação.

Nesse ponto, destaca-se que, diferente da hermenêutica, a análise de discurso não procura um sentido verdadeiro através de uma "chave" de interpretação, pois não há esta chave, há método, há construção de dispositivo teórico. "Não há uma verdade oculta atrás do texto. Há gestos de interpretação que o constituem e que o analista, com seu dispositivo, deve ser capaz de compreender." (ORLANDI, 2015, p. 24)

Giovani Forgiarini Aiub (2012, p. 65) esclarece:

> Cabe ressaltar, porém, que a Análise do Discurso não é uma ferramenta que desvenda o sentido verdadeiro de arquivos ocultos. Muito longe disso. A Análise do Discurso trata de "[...] desvendar os processos discursivos que levam à imposição como evidência, bem como o que esses mesmos processos deixam de fora" (MITTMANN, 2007, p. 154). Por isso que está em seu alicerce o gesto de leitura de arquivos a fim de constituir um corpus analítico.

Como consequência do gesto de interpretação, emerge do procedimento analítico a necessidade de fazer recortes no corpus de pesquisa, de modo a direcionar a análise para um ponto determinado. Segundo Orlandi (1984, p. 14), o *recorte* é uma unidade discursiva que, por sua vez, são fragmentos correlacionados de linguagem-e-situação. "Assim, um recorte é um fragmento da situação discursiva."

Para Orlandi (1984, p. 14), o recorte varia de acordo com o tipo de discurso, as condições de produção, o objetivo e o alcance da análise. Nesse sentido, "*o texto é o todo em que se organizam os recortes*. Esse todo tem compromisso com as tais condições de produção, com a situação discursiva." Os recortes, segundo a autora, são feitos na (e pela) situação de interlocução, aí compreendido um contexto (de interlocução) menos imediato: o da ideologia.

Por isso, ao construirmos o corpus de análise, segundo Flores (2014, p. 55), "estamos recortando, decidindo acerca de propriedades discursivas, isto é, marcando a posição discursiva que nos orienta para a construção desse corpus de investigação."

A partir de tais considerações, construímos o arquivo do corpus de pesquisa por meio da seleção de alguns artigos, científicos ou não, bem como *podcast*, que tratam da

divulgação do Projeto Victor e/ou crítica a seu respeito, ambos não apenas no ambiente acadêmico científico, mas jurídico como um todo, alcançando advogados, juízes, promotores de justiça, servidores do Poder Judiciário, etc.

É importante destacar desde já que os textos selecionados tiveram como premissa básica a apresentação do discurso de divulgação científica; contudo, os mesmos textos não necessariamente tem essa formação discursiva como dominante, ou seja, alguns tem como dominante a formação discursiva relativa ao discurso científico (tese de doutorado, artigo científico), em outros a formação discursiva dominante é relativa ao discurso jornalístico (*podcast*, notícia em site), porém todos são atravessados pela formação discursiva relativa ao discurso de divulgação científica, que nos casos que vamos analisar, representa uma formação discursiva não dominante, no complexo de formações discursivas. Desse modo, o ponto em comum é o discurso de divulgação científica, porém cada um deles, em razão da particularidade da sua formulação, circulação e finalidade, apresentará dominância de outros discursos (jornalístico, científico, jurídico) até porque, o discurso não é um monolito, mas é formado por um complexo de formações discursivas, o que explica sua natureza heterogênea.

Esclarecemos, ademais, que os textos selecionados são marcados por diferentes posições sujeito: cientista, advogado/a, pesquisador/a, divulgador/a do Projeto Victor, jornalista, e da mesma forma, a cada sequência discursiva, mais de uma posição sujeito pode fazer-se presente em razão das diferentes formações discursivas que constituem cada sujeito falante/escritor. Assim, poderemos ter um sujeito que marca sua posição no texto, predominantemente como pesquisador (em uma tese, por exemplo), e ocasionalmente marcar-se, no mesmo texto, na posição sujeito divulgador, ou na posição-sujeito advogado, etc, como veremos.

A composição do arquivo do corpus (gesto), portanto, levou em consideração não apenas os textos produzidos pelo coordenador do Projeto Victor, Prof. Dr. Fabiano Hartmann Peixoto, da UnB, de seus alunos e orientandos, que divulgam o projeto como um salto de inovação tecnológica dentro do campo jurídico e ressaltam a possibilidade de aproveitamento dos servidores judiciários em atividades mais complexas e celeridade processual, mas também a visão crítica de juristas preocupados com a transparência dos algoritmos, com a informação prévia do jurisdicionado sobre a adoção da IA em seu processo, sobre o direito de revisão das decisões automatizadas, bem como o efetivo acesso à justiça amparado no direito subjetivo de acesso aos juízes humanos.

Em ordem cronológica de publicação, o arquivo é composto pelos seguintes textos: a) Projeto Victor: perspectivas de aplicação da inteligência artificial ao direito, de Maia Filho e

Junquilho (2018); b) IA Projeto Victor, conversa (via *podcast*) com o Prof. Dr. Fabiano Hartmann, no IAJUSTEAM, grupo de estudos de IA e Direito, coordenado pelo Prof. Dr. Fausto Santos de Moraes (2019); c) Inteligência Artificial para o rastreamento de ações com repercussão geral: o Projeto Victor e a realização do princípio da razoável duração do processo, de Pinto, Lima e Galvão (2020); d) Tese estuda projeto pioneiro da UnB de inteligência artificial para o Poder Judiciário, de Pires (2020); e) Projeto Victor: relato do desenvolvimento da Inteligência Artificial na Repercussão Geral do Supremo Tribunal Federal, de Peixoto (2020); f) IA no Judiciário deve garantir ética, transparência e responsabilidade (Conjur, 2020); g) Princípio da Transparência Algorítmica e Devido Processo Legal: um diálogo necessário para garantia do direito à explicação, de Souza (2020); h) Inteligência artificial na tomada de decisões judiciais: três premissas básicas, de Roque e Santos (2021); i) Manual de inteligência artificial no direito brasileiro, de Fernanda de Carvalho Lage, 2021.

Constituído o arquivo do corpus de pesquisa, passamos ao corpus de análise, que é composto por algumas sequências discursivas especialmente produtoras dos efeitos que buscamos mostrar, dada a inviabilidade de mostrar tais elementos em todo o material.

Nesse ponto, Solange Mittmann destaca em seu artigo "Discurso e texto: na pista de uma metodologia de análise" que:

> Assim é que efetuamos nosso gesto de recortar sequências discursivas, isto é, nosso gesto arqueológico de relacionar sequências linguísticas, formando matrizes parafrásticas, definindo a Formação Discursiva (FD) dominante, delimitando suas fronteiras (...)

Faremos, na posição-sujeito analista, o gesto de leitura do arquivo por meio desse olhar, recortando as sequências discursivas de seu corpus e relacionando-as às matrizes parafrásticas a partir do aparato teórico-metodológico da análise do discurso, bem como fazendo relações com uma formação discursiva de referência, pensando as posições-sujeito, delimitando suas fronteiras (AIUB, 2012, p. 70 e 75).

Giovani Aiub (2012, p. 75-76) deixa inda mais clara a importância do recorte das sequências discursivas na análise:

> E este ir e vir deve ocorrer também com a leitura do arquivo. O analista deve tentar compreender os processos de constituição dos sentidos, para que possa recortar as sequências discursivas para análise. Este trabalho de ir e vir não cessa até que o próprio analista ponha um ponto (que não é final) no processo de análise. O analista de discurso deve compreender, antes de mais nada, que a leitura que ele faz do arquivo é uma possível entre outras e que seu trabalho aparece quando ele consegue compreender os processos discursivos, ou seja, os efeitos de sentido. O trabalho do analista não se esgota, mas é reflexo de um trabalho sócio-histórico de análise.

Considerando tais premissas, faremos a análise do discurso de divulgação do Projeto Victor, através de sequências retiradas do arquivo construído, em que se percebe, lendo e relendo, repetições, modos de abordagem mais ou menos comuns, uma historicidade recorrente. Nos textos selecionados, encontramos alguns elementos "chave": pré-construídos, memória, interlocutores convocados e, do complexo de formações discursivas, aquelas que são mais determinantes dos sentidos.

Dimitri Dimoulis (2020, p. 150) reforça tal pensamento quando diz que o discurso jurídico utiliza modos de expressão técnicos, concisos, repetitivos e "secos" a fim de evitar problemas das linguagens naturais. Segundo ele, quanto mais rigorosa (absoluta) for a linguagem jurídica, menor será o espaço deixado à polissemia, à ambiguidade sintática, à vagueza e às avaliações subjetivas. Assim, por exemplo, o legislador que deseja regulamentar a taxa de juros pode estabelecer um valor (10% ao ano), remeter a índices econômicos (inferior ao dobro da inflação) ou fazer indicação vaga (taxa de juros razoável). Depende do nível de controle que o legislador escolherá, determinado por si ou de acordo com o mercado.

Verifica-se, portanto, que o discurso jurídico é composto por saberes que estão no interdiscurso, com origem em uma formação histórica, além de uma formação discursiva, e é composto por diferentes discursividades, nas quais tais saberes estão materializados. Na realidade, toda a sociedade, justamente pela via das formas históricas do religioso e do jurídico, está sustentada na admissão de que existe uma verdade incontestável em alguma instância, e que alguém pode garanti-la, que pode ser Deus, a lei, a ciência, etc. O rigor metodológico do discurso científico garante, como acima dito, esse mesmo efeito de verdade (científica), um resultado indubitável, por meio de processos diferentes.

É diante de todo esse contexto, portanto, que envolve a historicidade e a memória dos discursos jurídico e científico, que o discurso de divulgação científica se constitui, produzindo sentidos do que vem a ser a inteligência artificial (IA) e, particularmente, a IA no direito.

4.1 PRIMEIRO RECORTE: A NEGAÇÃO

A fim de demonstrar a construção desse sentido, num primeiro recorte, trazem-se algumas sequências discursivas, extraídas do discurso de divulgação científica do PV, em que estão presentes uma negação, uma vez que toda negação pressupõe uma afirmação como sentido pré-construído, tais como:

> (SD 01) [...] **não é** o algoritmo quem decide [...] (PEIXOTO, 2019, 22’00’’a 22’20’’, *podcast*)

> (SD 02) [...] o objetivo do projeto **não é** que o algoritmo tome a decisão final acerca da repercussão geral. (MAIA FILHO e JUNQUILHO, 2018, p. 226)

Assim, quando o coordenador do Projeto Victor diz que “não é o algoritmo quem decide...” (PEIXOTO, 2019, 22’00’’a 22’20’’, *podcast*), ele retoma um enunciado de outro discurso que está pré-construído na sua fala (o algoritmo decide), o que remete ao esquecimento nº 2 de Michel Pêcheux (2014, p. 161), segundo o qual:

> [...] todo o sujeito-falante ‘seleciona’ no interior da formação discursiva que o domina, isto é, no sistema de enunciados, formas e sequências que nela se encontram em forma de paráfrase – *um enunciado, forma ou sequência, e não um outro, que, no entanto, está no campo daquilo que poderia reformulá-lo na formação discursiva considerada*. (com grifo no original)

O funcionamento que é descrito por Pêcheux (2014, p. 161) como o esquecimento nº 2, pode ser compreendido a partir do funcionamento que Eni Orlandi (2001, p. 9) descreve ao tratar das projeções imaginárias realizadas pelo autor. Quando o autor textualiza, ele dirige o seu texto a um leitor imaginário, um leitor ideal, numa projeção imaginária. Todo o texto, portanto, já tem embutido um leitor ideal com o qual leitor real se relacionará. Quanto maior a identidade entre o leitor ideal e o leitor real, mais legível será o texto, quanto menor, menos legível. Nessa perspectiva, afirma Eni Orlandi (2012, p. 10) que:

> Há um leitor virtual inscrito no texto. Um leitor que é constituído no próprio ato da escrita. Em termos que denominamos “formações imaginárias” em análise de discurso, trata-se aqui do leitor imaginário, aquele que o autor imagina (destina) para seu texto e para quem ele se dirige. (...)
> Assim, quando o leitor real, aquele que lê o texto, se apropria do mesmo, já encontra um leitor aí constituído com o qual ele tem de se relacionar necessariamente.

Desse modo, quando o autor textualiza o seu pensamento, projeta de forma imaginária suas palavras para um leitor ideal, a quem ele direciona seu texto.

O esquecimento nº 2 de Pêcheux (2014, p. 161), portanto, tem relação com essa seleção que faz o autor (sujeito falante) no interior da formação discursiva que o domina. Por que o autor seleciona? Como ele seleciona? Porque ele tem um interlocutor, um leitor imaginado. É esse o parâmetro que faz o autor selecionar um enunciado e não outro, uma sequência discursiva e não outra. Ele textualiza para alguém determinado.

No recorte realizado nesta pesquisa, portanto, observa-se que quando o sujeito pesquisador afirma que a inteligência artificial (PV) criada para uso na mais alta Corte brasileira (STF) não decide, ele está falando com seu interlocutor imaginado, aquele que já lhe interrogou

sobre a inviabilidade da máquina decidir algo que somente a consciência humana teria legitimidade para realizar.

É nesse sentido que em notícia veiculada pelo Supremo Tribunal Federal, em 30 de maio de 2018, afirmou-se que:

> (SD 03) O objetivo inicial é aumentar a velocidade de tramitação dos processos por meio da utilização da tecnologia para auxiliar o trabalho do Supremo Tribunal. A máquina **não** decide, **não** julga, isso é atividade humana. Está sendo treinado para atuar em camadas de organização dos processos para aumentar a eficiência e velocidade de avaliação judicial. (STF, 2018)

O sujeito divulgador, portanto, projeta imaginariamente seu texto ao leitor ideal, ou seja, àqueles sujeitos do discurso jurídico que contestam o uso da inteligência artificial sem critérios e afirmam, por exemplo, que "decisões tomadas exclusivamente por robôs devem ser de alguma forma submetidas à revisão humana." (ROQUE e SANTOS, 2021, p. 74)

O coordenador do Projeto Victor, em artigo científico publicado em 2020 (dois anos após o lançamento do projeto: 2018), Fabiano Hartmann Peixoto (2020, p. 2) procura novamente dar respostas a este interlocutor imaginário que, desde o início, questiona a real função do Projeto Victor:

> (SD 04) Logo no dia 01 de junho, o Estadão[3] apresentou o Victor de uma forma chamativa, como "o 12º Ministro do STF", dando a dimensão – inclusive política, o caráter inovador e a importância da pesquisa para o cenário da IA no Direito. As notícias, desde então, sugerem ou especulam alguns fatos sobre o projeto Victor e, esse artigo (vencidas as etapas que consumiam todas as energias da equipe), tem a função justamente de relatar o desenvolvimento da pesquisa, que (demonstrando acertos e verificação de oportunidades), foi renovada e estendida ainda pelo ano de 2020. Portanto, o projeto Victor segue, inclusive com escopo ampliado em relação ao plano original.

Dois anos após o início da pesquisa, além de desfazer contradições divulgadas, inclusive pelo discurso jornalístico de que o PV (um algoritmo) seria o 12º Ministro (são 11 Ministros que compõe o Supremo Tribunal Federal), era necessário produzir consenso na comunidade jurídica por meio de números que demonstravam a acurácia da IA, a retirada de trabalhos enfadonhos dos servidores e, com isso, a celeridade dos julgamentos. Em outras palavras, mostrar trabalho e resultados positivos. E desse modo, confirmar outros efeitos de sentido já divulgados em outras oportunidades, tais como, afirmar que o Projeto Victor:

> (SD 05) [....] é uma solução de fluxo com um indicador (...), ele é assim uma ferramenta de apoio, quem permanece decidindo é o servidor. (PEIXOTO, 2019, 22'54'', *podcast*).

---

[3] VICTOR, o 12º ministro do STF. *In:* ESTADÃO. 01 jun. 2018. Disponível em https://politica.estadao.com.br/blogs/fausto-macedo/victor-o-12-o-ministro-do-supremo/. Acesso em: 21 mar 2021.

> (SD 06) [...] isso pode ser assim, parecer um detalhe pequeno, mas acredito eu que responde muita coisa. (PEIXOTO, 2019, 22’56’’, *podcast*).

> (SD 07) [...] o Victor, ele **não é** um ditador de tendências, ele é um instrumento, efetivamente é um instrumento né! (PEIXOTO, 2019, 26’10’’, *podcast*).

Ou seja, Peixoto (2019, 22’57’’, *podcast*) reconhece que há muita crítica no ambiente jurídico (sentidos pré-construídos) quanto ao uso da IA, especialmente quando se tratam das decisões judiciais, que ao fim de todo o trâmite processual - que pode levar em média cincos anos se levado até o Supremo Tribunal Federal - decidem as vidas dos cidadãos brasileiros. Assim, quando ele diz que o PV não é uma solução de decisão, que não é o algoritmo quem decide, mas é uma ferramenta de apoio, ele procura restabelecer um consenso no discurso jurídico.

Além disso, quando Peixoto (2019, 22’57’’, *podcast*) afirma que “quem permanece decidindo é o servidor”, ele recupera um enunciado de outro sentido pré-construído, ou seja, a preocupação com o futuro do profissional do direito e das habilidades que permanecerão necessárias com o uso da IA, não apenas por servidores (juízes e servidores do Poder Judiciário), mas advogados, promotores de justiça e inclusive acadêmicos de direito. Aliás, na mesma ocasião, Peixoto afirma que o Projeto Victor:

> (SD 8) [...] “**não vai** provocar demissão, porque esse era um dos compromissos éticos da UnB, **não quer** gerar algo para causar o mal para as pessoas né” (PEIXOTO, 2019, 23’55’’, *podcast*).

É interessante destacar, aqui, o atravessamento do discurso capitalista sobre o discurso científico na fala acima, dirigida a um público de acadêmicos de direito, servidores do Poder Judiciário e advogados, na informalidade de um *podcast,* Diante de tais condições de produção, o sujeito divulgador procura deixar claro que o PV não provocará demissões, especialmente num momento (2019) pós-reforma trabalhista (Lei n. 13.467, de 13 de julho de 2017), em que a sociedade brasileira se preocupava com a manutenção de seus empregos.

Ao produzir o primeiro trabalho científico sobre a IA e direito na Universidade de Brasília, a tese de doutorado intitulada “A inteligência artificial na repercussão geral: análise e proposições da vanguarda de inovação tecnológica no Poder Judiciário brasileiro” (Pires, 2020), posteriormente publicada com o título Manual de Inteligência Artificial no Direito Brasileiro, Fernanda Carvalho Lage, orientada por Fabiano Hartmann Peixoto, destaca que:

> (SD 9) O papel do *machine learning* mostra-se, portanto, exclusivamente de apoio procedimental, sem carga decisória. Ele **não é** nem inicial (por se tratar de um exame em

> sede de recurso extraordinário), **nem** definitivo (vez que a máquina é apenas um apoio à decisão do ministro). (LAGE, 2021, p. 314)

Constroem-se, portanto, precipuamente da Universidade de Brasília para a comunidade jurídica brasileira, efeitos de sentidos sobre a IA no direito, que procuram esclarecer que o PV não interferirá na decisão do ministro, isto é, do juiz humano, sendo esta apenas uma ferramenta de apoio. Nesse caso o referido discurso de divulgação científica tem como alvo, como interlocutor, não somente o sujeito do discurso jurídico, mas também a sociedade civil, que tem que ser convencida que a IA pode e deve ser usada no direito, que ela vem para auxiliar nos trabalhos, acelerar processos e julgamentos, e não para ferir direitos e garantias fundamentais constitucionais, tais como os princípios[4] do juiz natural (art. 5º, XXXVII e LIII da CFRB/88), personificado em juízes aprovados em concurso público; e, do acesso à justiça (art. 5º, XXXV da CFRB/88), que pressupõe o acesso ao Poder Judiciário. (ROQUE e SANTOS, 2021, p. 71)

Assim, respeitar o princípio do juiz natural é dizer que temos o direito e a garantia de somente sermos processados e julgados por um tribunal ou juiz imparciais, que já estão definidos muito antes de o fato ocorrer, por meio de regras processuais de competência. (MASSON, 2021, p. 270-271)

Portanto, quando Roque e Santos (2021, p. 71) questionam o respeito ao princípio do juiz natural, consideram que o uso da IA em decisões judiciais estaria negando o acesso à decisão por um juiz humano e pré-determinado:

> (SD 10) [...] seria inconstitucional a tomada de decisões exclusivamente por robôs, sem que suas decisões sejam de alguma forma submetidas à revisão humana, sendo assegurado pela Carta Magna o direito público subjetivo de acesso aos juízes. (ROQUE e SANTOS, 2021, p. 71)

Quando os autores acima destacam a importância do princípio do acesso à justiça, segundo o qual todos nós temos o direito de obter do Poder Judiciário uma resposta aos pedidos a ele dirigidos, defendem que a garantia fundamental de acesso à justiça não se resume apenas ao direito subjetivo de se obter uma decisão judicial em caso de lesão ou ameaça a direito, mas que tal informação provenha de um juiz (que se presume humano se conjugado este princípio

[4] Princípio, para o direito, tem o sentido de início, origem, nascente. Pode ser compreendido como normas que ordenam a realização de algo na maior medida possível, dentro das possibilidades jurídicas e fáticas. (ALEXY, 1993, p. 48 e 172) Ou ainda, pode ser caracterizado como um eixo central que orienta determinado conjunto de ideias e ações. (CASTAGNA, 2019, p. 36) Desse modo, na interpretação das normas jurídicas, os princípios são norteadores e devem ser respeitados indistintamente.

com aquele do juiz natural), que terá examinado o processo e apresentado o fundamento adequado para a sua decisão (GONÇALVES, 2016, p. 65-66).

Como se verifica, portanto, os questionamentos acima, a respeito do uso da IA no direito, fragilizam o tradicional efeito de verdade absoluta, produzido pelo discurso jurídico, que sustenta a noção de sujeito-de-direito (HAROCHE, 1992, p. 179), e desse modo, colocam em xeque a transparência (publicidade dos atos) que se espera seja garantida nas decisões judiciais. Entretanto, se analisada a transparência sob o ponto de vista do efeito, ver-se-á que não existe, de fato, transparência no sentido, pois o sentido é sempre opaco, passível de equívoco, não evidente (ORLANDI, 2020, p. 9).

O próximo recorte, portanto, a ser desenvolvido na presente pesquisa refere-se à questão da transparência.

4.2 SEGUNDO RECORTE: A TRANSPARÊNCIA

Como vimos anteriormente, a noção de *discurso* para Pêcheux (2014, p. 81) está relacionada com a produção de efeitos de sentidos entre interlocutores, e não na simples transmissão de informação, ou seja, tratam-se de efeitos de sentidos relacionados às diferentes posições que os sujeitos podem ocupar nos discursos. Assim, Pêcheux exemplifica, no interior da esfera de produção econômica, os lugares de "patrão", funcionário de repartição, contramestre e operário, os quais são marcados por propriedades diferenciais determináveis. Referidas propriedades estarão significando sempre, na fala dos sujeitos nessas posições, independente de sua intenção. E é justamente essa espessura histórica, a materialidade dos discursos, que permite a interpretação. Não poderíamos interpretar uma fala, qualquer que fosse, sem parâmetros sociais e históricos. Essa materialidade funciona como memória, sem a qual os sentidos não se constituem e, ao mesmo tempo, confere a eles, em certa medida, sua determinação. Interessa salientar que essa determinação não é percebida pelos sujeitos, senão como apagamento. Pêcheux (2014, p. 121) propõe esse como sendo o efeito ideológico elementar, que coloca o sujeito imaginariamente em uma eterna origem do sentido, em plena transparência, esquecido daquilo que o determina.

Eni Orlandi (2015, p. 19) traz a noção de discurso ao dizer que não se trata de transmissão de informação. Segundo a autora "a linguagem pode comunicar e não comunicar", pois no funcionamento da linguagem, que põe em relação sujeitos e sentidos afetados pela língua e pela história, há um complexo processo de constituição de sujeitos e produção de sentidos. (ORLANDI, 2015, p. 19)

Os sentidos são produzidos, portanto, de acordo com a constituição de cada sujeito, afetado pela língua e pela história. Mas não apenas. Segundo Orlandi (2015, p. 46), para que haja sentido, a ideologia aparece como efeito de relação necessária do sujeito com a língua e com a história. Uma relação de apagamento. Afirma a autora que:

> [...] a interpelação do indivíduo em sujeito pela ideologia traz necessariamente o apagamento da inscrição da língua na história para que ela signifique produzindo o efeito de evidência do sentido (o sentido-lá) e a impressão do sujeito ser a origem do que diz. (ORLANDI, 2015, p. 46)

Os efeitos mencionados por Eni Orlandi (2015, p. 46) - evidência de sentido e impressão de o sujeito ser a origem do que diz – demonstram a ilusão que temos de criação de algo completamente inédito e de uma língua vazia, uma estrutura que serviria a qualquer conteúdo. No entanto, a língua não funciona sem o discurso que lhe atribui sentidos, a cada tomada da palavra. Desse modo, não é possível mobilizar a língua para expressar-se com clareza. A linguagem não é evidente, não é transparente, mas opaca, espessa, obscura. Segundo ela, são:

> [...] efeitos que trabalham, ambos, a ilusão da transparência na linguagem. No entanto nem a linguagem, nem os sentidos nem os sujeitos são transparentes: eles têm sua materialidade e se constituem em processos em que a língua, a história e a ideologia concorrem conjuntamente. (ORLANDI, 2015, p. 46)

Na medida em que a transparência é inatingível na linguagem, que não existe transparência no sentido, porque o sentido é sempre opaco (denso e parcialmente apagado), o que é passível de se produzir, é apenas um efeito de transparência (ORLANDI, 2015, p. 59), mas não a transparência em sua plenitude.

Diante disso, seleciono mais algumas sequências discursivas (SDs) propostas para a análise, e procuro, conforme Orlandi (2015, p. 59), atravessar o efeito de transparência da linguagem, da literalidade do sentido e da onipotência do sujeito, para observar que a transparência, inclusive no discurso jurídico, não se verifica.

Sabemos, entretanto, que outras formas de interpretação tem a transparência como um fim. Esse é o caso da hermenêutica. Por outro lado, a ciência, de maneira geral, e a ciência da computação, de maneira específica, tem contornado o problema da transparência, relativizando suas verdades, de acordo com suas premissas. É justamente essa discussão que veremos aqui entre um discurso de divulgação científica e o discurso jurídico. O primeiro leva em conta uma verdade relativa aos saberes científicos que ele divulga, o segundo procura garantir a verdade e a transparência, como resultantes da correta interpretação.

Ao relatar o contexto e desenvolvimento do Projeto Victor, em artigo escrito em 2020, esclarecendo os acontecimentos desde o início da pesquisa em 2018, Fabiano Hartmann Peixoto destaca a eficiência do projeto ao tratar da diminuição da quantidade de processos no Supremo Tribunal Federal:

> (SD 11) Um fator crítico da pesquisa é a contribuição para a redução do acervo do STF, uma das preocupações na gestão estratégica da Corte. Os parâmetros de volume devem ser observados em uma perspectiva de acervo e recebimento. O dado que se destaca é o do contexto do início do ano de 2018, em que o STF estava em um cenário de aumento no recebimento de processos na ordem de 14,74%, com 103.650 processos recebidos pela Corte em 2017. (PEIXOTO, 2020, p. 12-13)

Para sustentar sua afirmação, de modo a comprovar a eficiência da inteligência artificial desenvolvida, Peixoto (2020, p. 13) apresenta dados estatísticos que demonstram a evolução do recebimento de processos e a evolução do acervo, ambos no STF, evidenciando a sua diminuição de 2018 para 2020:

(SD 12):

Tabela 4 – Estatísticas combinadas

Evolução do recebimento de processos do STF	Ano	Total	Percentual
	2015	93.477	
	2016	90.331	- 3,36%
	2017	**103.650**	**+14,74%**
	2018	101.497	-2,07%
	2019	93.197	-8,17%

Fonte: Supremo Tribunal Federal

Tabela 5 – Estatísticas combinadas

Evolução do Acervo do STF	Ano	total	Redução percentual
	2018	38.657	19,08 %
	2019	31.279	19,2 %
	2020	26.046	16,7 %

Fonte: Supremo Tribunal Federal

No mesmo sentido, Mamede Said Maia Filho[5], integrante do grupo de pesquisa e desenvolvimento do Projeto Victor, e Tainá Aguiar Junquilho[6], bolsista da FINATEC do

[5] http://lattes.cnpq.br/6611037107398589

[6] http://lattes.cnpq.br/5848504606151120

mesmo projeto, ambos Professores da UnB, tratam da eficiência da IA desenvolvida em artigo científico publicado no último trimestre de 2018. No trecho abaixo, citam Eduardo Silva Toledo[7], Diretor-Geral do STF de 2016 a 2020, e Jamie Backer[8], Professora e Diretora Associada da Texas Tech University School of Law:

> (SD 13) A experiência do Projeto Victor traz luz às perspectivas que a IA e a tecnologia podem gerar, quando aplicadas ao Poder Judiciário. Dentre os prognósticos do que pode ocorrer, tendo em conta as pesquisas que estão em curso, é de se ressaltar: *a)* a redução no tempo de tramitação de processos, em virtude da automação de procedimentos técnicos, o que fortalece, inclusive, a concretização do princípio da eficiência administrativa (TOLEDO, 2018); *b)* o desenvolvimento de tecnologias e pesquisas genuinamente brasileiras, que levem em conta as particularidades do nosso congestionado sistema judicial; *c)* o incremento da agilidade e eficácia das ferramentas de consulta processual e jurisprudencial, o que gera também economia de tempo, precisão e coerência institucional (BAKER, 2018); *d)* o tratamento isonômico das questões apresentadas ao Judiciário, que torna mais eficazes os princípios do contraditório, da ampla defesa e do livre acesso à justiça. (MAIA FILHO e JUNQUILHO, 2018, p. 230)

Como vemos, no discurso de divulgação científica há uma aproximação com a forma de interpretação da ciência da computação, no entanto, outra forma se coloca no horizonte da discussão, a partir do discurso jurídico, como vemos na seguinte SD, proferida por Andre Vasconcelos Roque e Lucas Braz Rodrigues dos Santos, ao submeter artigo científico, aceito e publicado na Revista Eletrônica de Direito Processual – REDP no primeiro quadrimestre de 2021:

> (SD 14) Frise-se que no ordenamento jurídico brasileiro, o princípio ético da transparência algorítmica revela-se como substrato do próprio princípio da publicidade (art. 5º, LX e 93, IX da Constituição e art. 8º do CPC). Se não há a devida transparência é impossível exercer controle – *accountability*[9] – sobre a adequada utilização da inteligência artificial. (ROQUE e SANTOS, 2021, p. 69)

Tais vozes, que no discurso de divulgação científica funcionam como pré-construídos, são relacionadas aos possíveis leitores dos artigos, obras, entrevistas, palestras, *podcasts,* etc, a quem se deve responder. Assim, no trabalho de textualização, o sujeito divulgador procura acalmar essas vozes, consensuar essa "fissura" que o dispositivo de inteligência artificial (Projeto Victor) está produzindo, e desse modo, busca estabelecer um outro efeito de sentido de transparência, que não a transparência relacionada à verdade.

---

[7]https://pesquisa.in.gov.br/imprensa/servlet/INPDFViewer?jornal=529&pagina=50&data=11/09/2020&captchafield=firstAccess

[8] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2978703

[9] Tradução livre: prestação de contas.

Os interlocutores levantam a necessidade de respeito ao princípio da transparência algorítmica e direito à explicação clara e precisa de como se deu o processo de tomada de decisão, quais dados foram utilizados e atribuídos.

Entretanto, ao discorrer sobre transparência algorítmica em sua obra Inteligência Artificial e Direito, Fabiano Hartmann Peixoto (2019, p. 73) questiona, por outras vias, o próprio conceito de transparência, e afirma que a simples exposição do algoritmo pode gerar uma ilusão de clareza:

> (SD 15) Daí, justifica-se, para o que se pode chamar de um problema *standard*[10]*,* igualmente uma solução *standard*: uma transparência algorítmica. De início, Desai aponta para o mal-entendido que pode se originar pela cultura mítica que se desenvolveu em torno dos algoritmos. A transparência, por sua vez, é um conceito poderoso. Contudo, a pura exposição algorítmica, a pura e simples disponibilidade de códigos-fonte ou auditorias irrefletidas podem gerar uma ilusão de clareza.

Ao colocar em dúvida o próprio conceito de transparência e afirmar que a exposição do funcionamento do algoritmo não garante a almejada transparência, o sujeito divulgador do Projeto Victor reconhece, de certa forma, que a transparência, conforme compreendida no discurso jurídico, não é possível no caso dos algoritmos, e assim, projeta imaginariamente seu texto ao leitor ideal (esquecimento n. 2 de Pêcheux - 2014, p. 161), ou seja, àqueles sujeitos do discurso jurídico que exigem transparência nas situações em que a IA está presente em decisões judiciais (conforme SD8), dando a eles as características do algoritmo, ou seja, apontando-o como uma estrutura básica da qual se pode produzir inúmeras e diferentes leituras.

Noutra oportunidade, em entrevista via *podcast,* o sujeito divulgador rebate algumas críticas e afirma, após negações para esclarecer o que o PV não faz (acima expostas), que a emissão de decisões com base no algoritmo feriria o próprio direito (desvio de finalidade):

> (SD 16) O Victor não está preocupado com o que os juízes almoçam ou tomam café, né; o Victor não vai fornecer elementos para se fazer relatório fático de decisão; há trabalhos sugerindo alguma coisa nesse sentido, de outros pesquisadores, nós temos todo respeito por isso né.....mas eu tenho muito claro, e todos que trabalham comigo também tem muito claro, nós temos um campo tão vasto de dificuldades tão mais simples, que **é quase que um desvio de finalidade, com o perdão da expressão, você querer começar a emitir decisões judiciais com base no algoritmo; tem muita coisa antes disso, todo mundo que trabalha com o direito sabe** (...) (PEIXOTO, 2019, 30’09’’, *podcast*) (grifo nosso)

Assim, a busca pela transparência, exigida pelos interlocutores do discurso de divulgação científica, exposta por exemplo por Roque e Santos (2021, p. 69, acima citado), perderia o sentido, na medida em que a decisão judicial continuaria a ser produzida pela inteligência humana, nos moldes conhecidos juridicamente.

---

[10] Tradução livre: padrão.

Entretanto, a transparência relativa, para a qual aponta o sujeito divulgador da IA, continua sendo alvo de críticas.

Segundo Lara Oliveira Souza (2020, p. 55), a compreensão do processo de tomada de decisões remete às possibilidades de contestação de decisões injustas, garantindo contraditório e ampla defesa:

> (SD 17) A automação introduziu um aspecto surpreendente: limita o papel do ser humano a confiar de maneira quase cega, em decisões conduzidas por computador. Diante desse cenário, o direito é posto em um campo perigoso no qual o uso de algoritmos poderá acarretar em decisões inexplicáveis. A transparência, auditabilidade e explicações das decisões automatizadas são essenciais para garantia do devido processo legal. (SOUZA, 2020, p. 53)

Diante da crítica acima, Fernanda de Carvalho Lage (2021, p. 314), pesquisadora e igualmente divulgadora do Projeto Victor, diz que:

> (SD 17) A crítica da substituição do homem pela máquina no exame dos pressupostos do recurso extraordinário, com o prejuízo para a ampla defesa e transparência, não procede. Por igual, não é realista o temor de vieses na decisão, por obra do aprendizado de máquina.
> (...)
> A cadeia decisória, com ou sem a IA, permanece inalterada, o que afasta o temor de ofensas à ampla defesa ou ao contraditório.

Nesse ponto, esclarece-se que os princípios do devido processo legal, ampla defesa e contraditório são previstos na Constituição Federal de 1988 (art. 5º, LIV e LV) e procuram garantir o direito à defesa da forma mais ampla possível a todos aqueles que buscam exercer os seus direitos. [11]

Embora divulgadora do Projeto Victor, Fernanda de Carvalho Lage (2021, p. 165) reconhece a ausência e importância da transparência:

> (SD 19) Com vistas a preservar os direitos fundamentais é preciso saber como uma tecnologia de inteligência artificial toma decisões, a fim de ter certeza de que confiar nela não resultará em uma violação. Para avaliar se deve haver responsabilidade por negligência para uma decisão baseada em IA, os Tribunais precisam ser informados de como a IA tomou sua decisão.

Assim, se para os interlocutores e sujeitos do discurso jurídico, a ausência de explicação clara e precisa de como se deu o processo de tomada de decisão pelo algoritmo

[11] O devido processo legal pode ser compreendido como um conjunto de garantias processuais, formais e materiais que deverão ser observadas para que o mesmo se concretize, tais como o contraditório (direito de o sujeito contradizer tudo o que fora apresentado no processo pela parte adversa) e a ampla defesa (direito de apresentar, no curso do processo, todos os meios lícitos que permitam ao sujeito provar o seu ponto de vista). (MASSON, 2021, p. 278)

prejudica a transparência e, desse modo, fere-se o devido processo legal, contraditório e ampla defesa (SOUZA, 2020, p. 53); para os sujeitos divulgadores do Projeto Victor, o uso da IA, ao contrário, torna a ampla defesa e o livre acesso à justiça "mais eficazes", pois confere tratamento isonômico das questões apresentadas ao Judiciário (MAIA FILHO e JUQUILHO, 2018, p. 230).

Em outras palavras, cada discurso (jurídico e de divulgação cientifica) confere sentidos diversos a garantias constitucionais que buscam a defesa do jurisdicionado: para um o direito à explicação clara e precisa (transparência) de como se deu o processo de tomada de decisão é fundamental para garantir o devido processo legal, contraditório e ampla defesa; para o outro, a tomada de decisões de forma isonômica a outros casos semelhantes, evitando-se decisões contraditórias, é o que garante o acesso à justiça e à ampla defesa do cidadão, optando por silenciar-se a respeito da transparência exigida pelos primeiros.

É interessante notar que a preocupação apontada pelos interlocutores do discurso de divulgação científica, ao mencionar a importância da transparência algorítmica, já foi objeto de discussão no Conselho Nacional de Justiça – CNJ, tendo sido publicada em 25/08/2020 a Resolução n. 332, que dispõe sobre a ética, transparência e governança na produção e uso da IA no Poder Judiciário. Na ocasião, Eunice Prado, juíza do Tribunal de Justiça de Pernambuco, e Tarciso Dal Maso, Professor do Centro Universitário de Brasília (UniCeub) e consultor legislativo do Senado Federal, não envolvidos diretamente no PV, mas com interesse no desenvolvimento da IA, defendem a regulamentação como meio de garantia à almejada transparência:

> (SD 20) Temos que observar a pessoa que vai utilizar o sistema inteligente e que tem direito ao seu controle. Ou seja, o uso da inteligência artificial vem para melhorar, jamais para restringir. (...) ao utilizar a IA para melhorar sua eficiência e transparência, o Poder Judiciário tem como desafios a comunicação, devendo informar ao usuário que ele está lidando com um agente virtual (PRADO, 2020)
>
> (SD 21) (ele) apontou o aspecto inclusivo como primordial para a utilização da inteligência artificial na busca de alternativas pontuais e aplicáveis na realidade prática do arcabouço jurídico brasileiro. O desafio tem sido enfrentado com a régua da garantia e promoção dos direitos fundamentais. E a IA é um vetor para garantir esses direitos fundamentais e da administração da justiça, com transparência e ética. (MASO, 2020)

Apontar para a normatização da exigida transparência na IA, estabelecendo inclusive o sentido de transparência no artigo 8º da Resolução n. 332/2020 do CNJ[12], é mais

[12] Art. 8º Para os efeitos da presente Resolução, transparência consiste em:
I – divulgação responsável, considerando a sensibilidade própria dos dados judiciais;
II – indicação dos objetivos e resultados pretendidos pelo uso do modelo de Inteligência Artificial;
III – documentação dos riscos identificados e indicação dos instrumentos de segurança da informação e controle para seu enfrentamento;

um modo de produzir um efeito de transparência e de acalmar as vozes que afirmam que a sua falta impede a garantia constitucional ao devido processo legal, à ampla defesa e ao contraditório.

O discurso de divulgação científica do Projeto Victor, portanto, passa não apenas a produzir um outro efeito de sentido de transparência, mas atravessa o discurso jurídico e o normatiza, por enquanto por meio de uma Resolução do CNJ, porém futuramente talvez o seja por meio de lei sobre a inteligência artificial no Brasil, se aprovado o Projeto de Lei n. 240, de 11 de fevereiro de 2020, de iniciativa do Deputado Federal Léo Moraes, de Rondônia. Porém, a dificuldade em criar uma definição para inteligência artificial (conceito opaco), enquanto requisito jurídico para a transparência, que poderia inclusive ser aplicada ao Projeto Victor, é apontado por Fernanda de Carvalho Lage, ao citar Miriam Caroline Buiten (2019), como um empecilho à criação de um regime regulatório para o uso da IA no Poder Judiciário:

> (SD 22) Um requisito jurídico de transparência para a inteligência artificial, e de fato para qualquer regime regulatório, seria sua definição. No entanto, ainda não parece haver qualquer definição amplamente aceita mesmo entre os especialistas da área. As várias definições de IA usadas na literatura podem ser úteis para entendê-la, mas são inadequadas como base para novas leis. Embora algumas incertezas possam ser inerentes às novas tecnologias, mas é problemático centrar leis e políticas em torno do conceito opaco dessa nova tecnologia (BUITEN, 2019, p. 45, *apud* LAGE, 2021, p. 161)

Contudo, ainda que viável a construção futura de uma definição de transparência, sob o ponto de vista da IA aplicada ao direito, isso não garante a almejada clareza como exige o sujeito do discurso jurídico, pois como dito, a língua não funciona sem o discurso que lhe atribui sentidos, não sendo possível mobilizá-la para expressar-se com clareza, pois a linguagem não é evidente, não é transparente, mas opaca, espessa, obscura, como já dissemos.

Desse modo, a crença do sujeito do discurso jurídico, que se inscreve na hermenêutica e aplica tal teoria, firmado na certeza de encontrar a verdade (transparência) por trás do texto normativo, ao fixar o real sentido e alcance da norma jurídica (MONTORO, 2008, p. 420), é o mesmo que espera a transparência no uso da inteligência artificial pelos Tribunais (em especial aqui, o STF), pois acredita que a transparência que existia não pode ser perdida, prejudicada, sob a justificativa de celeridade processual e outros benefícios da IA.

IV – possibilidade de identificação do motivo em caso de dano causado pela ferramenta de Inteligência Artificial;
V – apresentação dos mecanismos de auditoria e certificação de boas práticas;
VI – fornecimento de explicação satisfatória e passível de auditoria por autoridade humana quanto a qualquer proposta de decisão apresentada pelo modelo de Inteligência Artificial, especialmente quando essa for de natureza judicial.

Em ambas as situações, no entanto, seja na decisão proferida pelo uso da hermenêutica jurídica tradicional, por meio da inteligência humana, seja pela decisão apoiada na inteligência artificial, cujo mecanismo não é revelado, não é público, o que se tem é tão somente um efeito de transparência e não a transparência em sua plenitude, pois esta é inatingível tanto na materialidade linguística, quanto na materialidade digital, na medida em que o sentido é sempre opaco (ORLANDI, 2015, p. 59).

Atento a tal realidade e críticas, o discurso de divulgação científica do Projeto Victor, por meio de seus pesquisadores, reconhece a limitação em divulgar os mecanismos usados pela IA do PV, pois isso permitiria ao sujeito que apresenta seu recurso ao STF, apresentar suas razões com base nos termos do algoritmo e, com isso, forçar a admissão de seu recurso:

> (SD 23) A publicidade dos mecanismos de *machine learning* encontra óbvia limitação. Não é possível indicar termos exatos utilizados no aprendizado da máquina, nem aquelas que serão por ela, afinal utilizados, pois isso permitiria a qualquer recorrente simplesmente agregar à sua peça aqueles termos constantes do algoritmo, forçando, artificialmente, a admissão de recursos. (LAGE, 2021, p. 322)

Assim, o sujeito divulgador do PV propõe a ampliação da discussão a respeito do emprego da IA, e sugere que "algumas medidas poderiam ser adotadas para obviar a crítica da obscuridade, complexidade ou confiabilidade" (LAGE, 2021, p. 323):

> (SD 24) A solução é propiciar um teste público da eficiência e confiabilidade do sistema de IA, nos moldes do que vem sendo realizado pelo Tribunal Superior Eleitoral (TSE) com relação à segurança da urna e da votação eletrônica. A Corte Eleitoral promove um teste publico de segurança, bem como a apresentação do código fonte do software que utiliza nas urnas, para fiscalização da Ordem dos Advogados do Brasil e do Ministério Público. (LAGE, 2021, p. 323) (com grifo no original)

As mesmas medidas acima citadas como usadas pelo TSE na segurança das urnas eletrônicas são propostas pelo sujeito divulgador como forma de deslocar o sentido de transparência e apaziguar as vozes contrárias ao uso da IA no direito, embora o mesmo sujeito afirme que "o Victor não decide", que "é apenas uma ferramenta de apoio à decisão". Assim, segundo Fernanda Lage (2021, p. 323), poderiam ampliar a transparência do emprego da IA: a apresentação e auditoria de códigos-fonte pelo STF à OAB e MP, antes de colocar o software da IA em prática, a fim de solicitar melhorias, resolver dúvidas ou conversar com os programadores, bem como a realização de um teste público simulado, comparando-se os dados obtidos com o exame humano dos mesmos casos, disponibilizando-os aos interessados em geral e dando-se a devida publicidade.

Verifica-se, portanto, a construção de um novo efeito de sentido de transparência no direito, negociado entre o discurso jurídico, que o defende com fundamento na hermenêutica jurídica, que busca o sentido e alcance da norma; e, o discurso de divulgação científica, que o apresenta por outras vias, buscando relativizar o efeito de transparência, aparentemente perdido com a IA, por meio da publicização e auditoria dos códigos-fonte da IA por instituições que desempenham funções essenciais à justiça (OAB e MP), e propondo a realização de testes públicos que simulem o uso da IA, aberto aos interessados em geral.

Por fim, embora a transparência não se concretize no direito, com ou sem IA, há um efeito de transparência cujo sentido tem sido alvo de profundas críticas quando associado ao uso da IA, principalmente em decisões do STF e demais Tribunais. Desse modo, o que se constata neste início da década de 2020, é uma verdadeira batalha e negociação a respeito do efeito de sentido de transparência, de maneira que não fique exposta a contradição entre interpretação algorítmica e interpretação hermenêutica jurídica.

Sabe-se, no entanto, que não seria possível a introdução da IA nas práticas jurídicas, se a verdade e a transparência, resultantes da interpretação hermenêutica, de fato já não fossem, em alguma medida, efeitos. E é justamente isso que está na sustentação do discurso de divulgação do projeto de IA, mas de uma forma um tanto velada, evitando expor a contradição com o discurso jurídico, no sentido de não admitir que a transparência que os interlocutores perseguem nunca existiu, e que na realidade sempre foi uma questão de legitimação política de novas práticas que procuram se inserir no direito. É o que já ocorreu, por exemplo, com a virtualização dos processos antes impressos, iniciada em 2006[13] e hoje legitimada pelo discurso jurídico. Porém, tornar público o referido embate, inviabilizaria a própria aceitação da IA no ambiente jurídico, razão porque o discurso de divulgação científica dissimula a existência desta contradição, justamente para permitir a entrada e a legitimação da IA no direito, e para isso afirma que se trata de mais uma ferramenta tecnológica útil aos Tribunais, aos atores do direito, aos servidores e aos jurisdicionados, permitindo, com base em dados estatísticos, maior agilidade processual e o uso da inteligência humana em atividades mais complexas.

Finalmente, a derradeira questão que nós podemos colocar, e que talvez seja objeto de análise futura, é se a apresentação e auditoria de códigos-fonte pelo STF à OAB e MP, antes de colocar o software da IA em prática, e a realização de um teste público simulado, comparando-se os dados obtidos com o exame humano dos mesmos casos, dando-se a devida publicidade, medidas essas sugeridas pelo sujeito divulgador (LAGE, 2021, p. 323) como forma

[13] Lei n. 11.419, de 19 de dezembro de 2006, que dispõe sobre a informatização do processo judicial.

de ampliar a transparência do emprego da IA, efetivamente solucionariam o problema da verdade e da transparência, ou se isso constitui apenas mais um passo para formular-se um discurso, uma historicidade, uma política, uma legitimidade que produzirá e garantirá um efeito de verdade e transparência, tal como já o é na hermenêutica praticada pelo discurso jurídico, desde a sua origem.

É com tal interrogação, ou talvez, provocação, que finalizamos provisoriamente a presente análise, num momento de transição histórica do discurso jurídico, atravessado pelo discurso de divulgação científica, com a finalidade de admissão da inteligência artificial em vários procedimentos executados originalmente pela inteligência humana.

5. CONCLUSÃO

A necessidade de um fecho, ainda que provisório, para a análise desenvolvida nesta pesquisa nos demanda, na posição-sujeito analista, a tarefa de buscar um arremate, um ponto final para este momento, amealhando os elementos que permitiram todo o caminho até aqui percorrido. Sem dúvida ainda há muito a pesquisar e analisar, pois como dito, o discurso jurídico passa por um momento de transição quando associado à inteligência artificial. Apesar disso, marcar esta historicidade academicamente, utilizando-se a análise do discurso como método para analisar este contexto, unir estas três áreas (Direito, IA e AD), é fundamental para projetar-se os passos que ainda estão por vir.

Desse modo, a pergunta discursiva que norteou a pesquisa ao longo de toda a caminhada neste último ano foi a seguinte: como se constitui, como se formula e como circula o discurso de divulgação do Projeto Victor, de forma a não entrar em contradição com a hermenêutica jurídica?

Partindo desse questionamento, iniciamos a análise por meio da definição do corpus de pesquisa, que já é um gesto de interpretação, e selecionamos materiais de divulgação do Projeto Victor, que começaram a ser divulgados a partir do seu nascimento, em 09/04/2018, com a parceria firmada entre o STF e a UnB, com a finalidade de analisar o texto dos processos que chegam no Supremo para classificá-los em algum tema reconhecido de repercussão geral. (PEIXOTO, 2020, p. 3)

Assim, construímos o arquivo do corpus de pesquisa por meio da seleção de alguns artigos, científicos ou não, bem como *podcasts*, que tratam da divulgação do Projeto Victor e/ou crítica a seu respeito, ambos não apenas no ambiente acadêmico científico, mas jurídico como um todo, alcançando advogados, juízes, promotores de justiça, servidores do Poder Judiciário, etc.

Esse arquivo, portanto, foi composto pelos seguintes textos: a) Projeto Victor: perspectivas de aplicação da inteligência artificial ao direito, de Maia Filho e Junquilho (2018); b) IA Projeto Victor, conversa (via *podcast*) com o Prof. Dr. Fabiano Hartmann, no IAJUSTEAM, grupo de estudos de IA e Direito, coordenado pelo Prof. Dr. Fausto Santos de Moraes (2019); c) Inteligência Artificial para o rastreamento de ações com repercussão geral: o Projeto Victor e a realização do princípio da razoável duração do processo, de Pinto, Lima e Galvão (2020); d) Tese estuda projeto pioneiro da UnB de inteligência artificial para o Poder Judiciário, de Pires (2020); e) Projeto Victor: relato do desenvolvimento da Inteligência

Artificial na Repercussão Geral do Supremo Tribunal Federal, de Peixoto (2020); f) IA no Judiciário deve garantir ética, transparência e responsabilidade (2020); g) Princípio da Transparência Algorítmica e Devido Processo Legal: um diálogo necessário para garantia do direito à explicação, de Souza (2020); h) Inteligência artificial na tomada de decisões judiciais: três premissas básicas, de Roque e Santos (2021); i) Manual de inteligência artificial no direito brasileiro, de Fernanda de Carvalho Lage, 2021.

A partir do corpus de pesquisa, selecionamos algumas sequências discursivas especialmente produtoras dos efeitos que buscamos mostrar a fim de compor o corpus de análise, que foi formado pelo recorte de vinte e quatro sequências discursivas (SDs), especialmente produtoras dos efeitos que buscamos mostrar, uma vez que seria inviável a apresentação dos elementos em todo o material do corpus de pesquisa.

Com o apoio do aparato teórico-metodológico da análise do discurso e na posição-sujeito analista, realizamos um gesto de leitura sobre o corpus de pesquisa e recortamos sequências em que se percebem repetições e uma historicidade recorrente: pré-construídos, memória, interlocutores convocados e atravessamentos de diferentes sentidos vindos do complexo de formações discursivas, cuja dominante é a do discurso de divulgação científica.

Admitindo-se que toda a sociedade, pela via das formas históricas dos discursos, notadamente do jurídico, se sustenta na existência de uma verdade incontestável em alguma instância, e que alguém pode garanti-la, que pode ser Deus, a lei, a ciência, etc., observamos que o discurso científico, assim como o jurídico, garante, por outros processos, esse mesmo efeito de verdade e um resultado indubitável.

É diante de todo esse contexto, portanto, que envolve a historicidade e a memória dos discursos jurídico e científico, que selecionamos o discurso de divulgação científica como norteador da análise, na medida em que é a partir dele e do seu papel mediador, que se produzem os sentidos do que vem a ser a inteligência artificial (IA) e, particularmente, a IA no e para o direito.

Iniciamos, num primeiro recorte, a partir de sequências discursivas extraídas do discurso de divulgação científica do Projeto Victor, aquelas em que estão presentes uma negação, uma vez que toda negação pressupõe uma afirmação como sentido pré-construído, tais como as SDs 01 e 02 acima, que afirmam, ao mencionar o Projeto Victor, que "não é o algoritmo quem decide" (PEIXOTO, 2019, 22'00''a 22'20'', *podcast*), que "o objetivo do projeto não é que o algoritmo tome a decisão final acerca da repercussão geral" (MAIA FILHO e JUNQUILHO, 2018, p. 226). Com isso observamos que tais negativas retomam um enunciado de outro discurso que está pré-construído ("o algoritmo decide"), o que remete ao esquecimento

nº 2 de Michel Pêcheux (2014, p. 161), segundo o qual, ao textualizar, o autor dirige o seu texto a um leitor imaginário, um leitor ideal, numa projeção imaginária (ORLANDI, 2001, p. 9).

Desse modo, a análise nos permitiu observar que o sujeito pesquisador-divulgador fala com o seu interlocutor imaginado quando afirma que a IA criada para a Suprema Corte brasileira não decide, e no mesmo sentido, responde às questões que já lhe foram lançadas a respeito da inviabilidade da máquina decidir algo que somente a consciência humana teria legitimidade para realizar.

Percebe-se, portanto, que há sentidos que estão sendo interditados pelos pesquisadores, em especial aqueles que tratam da contradição entre o discurso científico (aplicação da IA ao direito) e o discurso jurídico, e a partir disso procuram acalmar vozes divergentes, a exemplo de Roque e Santos (2021, p. 71), quando afirmam que o uso da IA em decisões judiciais estaria negando o acesso à decisão por um juiz humano e pré-determinado, direito constitucionalmente assegurado pelo princípio do juiz natural, conforme SD 10 citada.

Questionamentos como este acima, a respeito do uso da IA no direito, fragilizam o tradicional efeito de verdade absoluta, pelo discurso jurídico (HAROCHE, 1992, p. 179), o que nos levou a definir como segundo recorte, sequências que trouxessem a discussão sobre "transparência", colocada em xeque em função da falta de publicidade dos atos que são executados pelo algoritmo, ou seja, o advogado ou a própria parte interessada não tem acesso à informação sobre o uso da IA na decisão proferida pelo juízo, tampouco em que medida a tecnologia foi utilizada.

Portanto, com apoio em Orlandi (2020, p. 29), verificamos que se analisada a transparência sob o ponto de vista do efeito, não há, de fato, transparência no sentido, pois o sentido é sempre opaco, passível de equívoco, não evidente. Se o direito se materializa essencialmente pela linguagem, e se não é possível mobilizar a língua para expressar-se com clareza, uma vez que os sujeitos se constituem em processos em que a língua, a história e a ideologia concorrem conjuntamente (ORLANDI, 2015, p. 46), a transparência é inatingível na linguagem, seja falada, escrita, e mais ainda, quando processada por algoritmos.

Diante disso, o sujeito pesquisador e divulgador do Projeto Victor foca na eficiência do algoritmo, destaca a diminuição da quantidade de processos no Supremo Tribunal Federal e apresenta dados estatísticos dos anos de 2015 a 2020, o que pode ser visto nas SDs 11 e 12 (PEIXOTO, 2020, p. 12-13). No mesmo sentido, na SD 13 verifica-se que integrantes do grupo de pesquisa e desenvolvimento do Projeto Victor ressaltam o tratamento isonômico de questões apresentadas ao Judiciário, tornando eficazes os princípios do contraditório, ampla defesa e livre acesso à justiça (MAIA FILHO e JUNQUILHO, 2018, p.

230); e o Diretor-Geral do STF, à época do lançamento do projeto, enfatiza a redução no tempo de tramitação de processos, em virtude da automação de procedimentos técnicos, destacando a concretização do princípio da eficiência administrativa (TOLEDO, 2018). Logo, a reconhecida inviabilidade da transparência, alvo de críticas de interlocutores do discurso de divulgação científica (SD 14 - ROQUE e SANTOS, 2021, p. 69), leva os sujeitos divulgadores a dar destaque à eficiência da IA desenvolvida para o STF, acelerando a resposta que o Poder Judiciário deve prestar àqueles que o buscam.

Se de um lado, portanto, tem-se parte dos sujeitos do discurso jurídico levantando a necessidade de respeito à transparência do algoritmo e do direito à explicação clara e precisa de como se deu o processo de tomada de decisão com o seu uso e, desse modo, fere-se o devido processo legal, contraditório e ampla defesa (SOUZA, 2020, p. 53); de outro, vê-se o sujeito divulgador buscando acalmar essas vozes, consensuar essa "fissura" que o dispositivo de inteligência artificial (Projeto Victor) está produzindo, e desse modo, buscando estabelecer um outro efeito de sentido de transparência, que não a transparência como verdade.

Com base nessa realidade, uma pesquisa científica em nível de doutoramento na Universidade de Brasília propõe a ampliação da discussão sobre o emprego da IA e sugere que "algumas medidas poderiam ser adotadas para obviar a crítica da obscuridade, complexidade ou confiabilidade" (LAGE, 2021, p. 323), tal como se observou a partir da SD 24 transcrita. Para tanto, a pesquisadora e divulgadora do projeto sugere a apresentação e auditoria de códigos-fonte pelo STF à OAB e MP, antes de colocar o software da IA em prática a fim de solicitar melhorias, resolver dúvidas ou conversar com os programadores; e ainda, a realização de um teste público simulado, comparando-se os dados obtidos com o exame humano dos mesmos casos, disponibilizando-os aos interessados em geral e dando-se a questionada publicidade. Tal proposta, na realidade, é mais um passo para sedimentar um novo sentido de transparência, que permita o uso da IA sem necessariamente chocar-se com a hermenêutica jurídica.

Trata-se de uma verdadeira batalha em que sujeitos divulgadores e interlocutores negociam o deslocamento desse sentido, situação que somente pude compreender por meio do estudo da teoria da análise do discurso, que permitiu esta pesquisa. A textualização da presente dissertação, que apresentou as forças e disputas de sentidos em torno da admissão da IA no direito, apenas foi possível a partir do estudo e distinção dos discursos jurídico, científico e de divulgação científica.

Certamente que uma pesquisa como a presente, especialmente porque sustentada na análise do discurso, que se altera de acordo com as condições de produção, sujeitos

envolvidos, suas ideologias e historicidades, não se encerra por uma conclusão, porque um fecho final sempre poderá ser outro e é marcado pela provisoriedade.

De todo modo, registro a importância do percurso até aqui e deste fecho, ainda que provisório, pois me permitiu compreender muitos dos meandros e atravessamentos que estão presentes no discurso jurídico, antes incompreensíveis, ou sequer notados.

REFERÊNCIAS

ADINOLFI, Valéria Trigueiro Santos. **Discurso científico, poder e verdade**. Revista Aulas. N. 3 – dezembro 2006/março 2007. Organização Margareth Rago & Adilton Luís Martins. Disponível em https://www.ifch.unicamp.br/ojs/index.php/aulas/article/view/1940/1401 Acesso em 12 jun 2021

AIUB, Giovani Forgiarini. **Arquivo em análise do discurso: uma breve discussão sobre a trajetória teórico-metodológica do analista.** Revista Leitura. Maceió, n. 50, p. 61-82, jul/dez 2012. Disponível em https://www.seer.ufal.br/index.php/revistaleitura/article/view/1149/784 Acesso em 02/02/2021

ALEXY, Robert. **Teoria de los derechos fundamentales**. Madrid: Centro de Estudios Constitucionales, 1993.

ANDRADE, Mariana Dionísio de; PINTO, Eduardo Régis Girão de Castro; LIMA, Isabela Braga de; GALVÃO, Alex Renan de Sousa. **Inteligência Artificial para o rastreamento de ações com repercussão geral: o Projeto Victor e a realização do princípio da razoável duração do processo.** Revista Eletrônica de Direito Processual – REDP. Rio de Janeiro. Ano 14. Volume 21. Número 1. Janeiro a Abril de 2020. Disponível em < https://www.e-publicacoes.uerj.br/index.php/redp/article/view/42717/31777> Acesso em 07/01/2021

AZEVEDO, Bernardo. **Como a inteligência artificial está transformando a prática jurídica.** Disponível em < https://bernardodeazevedo.com/conteudos/como-a-inteligencia-artificial-esta-transformando-a-pratica-juridica/ > Acesso em 23 de Jul de 2020

AZEVEDO, Bernardo. **Código de ética global sobre inteligência artificial está prestes a se tornar realidade.** Disponível em <https://bernardodeazevedo.com/conteudos/codigo-de-etica-global-sobre-inteligencia-artificial-esta-prestes-a-se-tornar-realidade/?fbclid=IwAR02VyKJ_xS3RKAS-NnVGmvACkZkRKPC2Hu5PXDFJcfsNUeYwk_sGrPrxCg > Acesso em 23 de Jul de 2020

BENVENISTE, Émile. Vista d'olhos sobre o desenvolvimento da linguística. **Problemas de linguística geral.** Tradução de Maria da Glória Novak e Luiza Neri; revisão do Prof. Isaac Nicolau Salum. São Paulo: Nacional/USP, 1976.

BERSOT, Dayse Carias e LIMA, Jacqueline de Cassia Pinheiro. **Análise do discurso científico em um acervo de memória: caso do centro Pan-americano de febre aftosa OPAS/OMS.** Cadernos do CNLF, Vol. XVI, N. 4, t. 1 – Anais do XVI CNFL, 2012. Disponível em <http://www.filologia.org.br/xvi_cnlf/tomo_1/024.pdf > Acesso em 12 jun 2021.

BIDDLE, Sam; RIBEIRO, Paulo Victor e DIAS, Tatiana. **EXCLUSIVO: TikTok escondeu 'feios' e favelas para atrair novos usuários e censurou posts políticos.** Disponível em <https://theintercept.com/2020/03/16/tiktok-censurou-rostos-feios-e-favelas-para-atrair-novos-usuarios/> Acesso em 19 jun 2020

BOBBIO, Norberto. **Teoria da norma jurídica.** Traduzida por Fernando Pavan Baptista e Ariani Bueno Sudatti. Bauru: Edipro, 2001.

BRASIL. **Constituição da República Federativa do Brasil, de 05 de outubro de 1988**. Disponível em < http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm > Acesso em 14 Jul 2020.

BRASIL. **Lei n. 11.419, de 19 de dezembro de 2006.** Disponível em < http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2006/lei/l11419.htm> Acesso em 15 maio 2021.

BRASIL. **Lei n. 13.105, de 16 de março de 2015.** Código de Processo Civil. Disponível em < http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13105.htm > Acesso em 14 Jun 2020.

BRASIL. **Lei n. 13.467, de 13 de julho de 2017**. Altera a Consolidação das Leis do Trabalho (CLT). Disponível em < http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2017/lei/l13467.htm> Acesso em 26 Jul 2021.

BRASIL. **Lei n. 13.709, de 14 de agosto de 2018**. Disponível em < http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm > Acesso em 1º Ago 2020.

BRASIL. **Lei n. 14.010, de 10 de junho de 2020**. Disponível em <http://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2020/Lei/L14010.htm#art20 > Acesso em 1º Ago 2020.

BRASIL. **Lei n. 8.906, de 04 de julho de 1994**. Estatuto da Advocacia e a Ordem dos Advogados do Brasil (OAB). Disponível em < http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13105.htm> Acesso em 14 Jul 2020.

BRASIL. **Projeto de Lei n. 5051/2019, de 16 de setembro de 2019**. Estabelece os princípios para o uso da Inteligência Artificial no Brasil. Disponível em < https://www25.senado.leg.br/web/atividade/materias/-/materia/138790> Acesso em 03 Jun 2021.

BRASIL. **Projeto de Lei n. 21/2020, de 03 de fevereiro de 2020**. Cria o marco legal do desenvolvimento e uso da Inteligência Artificial (IA) pelo poder público, por empresas, entidades diversas e pessoas físicas. Disponível em <https://www.camara.leg.br/propostas-legislativas/2236340> Acesso em 03 Jun 2021.

BRASIL. **Projeto de Lei n. 240/2020, de 11 de fevereiro de 2020**. Cria a Lei da Inteligência Artificial, e dá outras providências. Disponível em < https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2236943 > Acesso em 03 Jun 2021.

BRASIL, Luciana Leão. Michel Pêcheux e a teoria da análise de discurso: desdobramentos importantes para a compreensão de uma tipologia discursiva. **Linguagem – Estudos e Pesquisas**. Catalão/GO, v. 15, n. 01, p. 171-182, jan/jun 2011.

BRASIL. Supremo Tribunal Federal. **Termo de Execução Descentralizada 01/2018**. Brasília: Supremo Tribunal Federal, 2018.

BUITEN, Miriam Caroline. ***Towards Intelligent Regulation of Artificial Intelligence***. *European Journal of Risk Regulation*. Vol 10:1, 2019, p. 41-59. Disponível em < https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/towards-intelligent-regulation-of-artificial-intelligence/AF1AD1940B70DB88D2B24202EE933F1B > Acesso em 1º Maio 2021.

CEPEJ, Comissão Europeia para a Eficiência da Justiça. **Carta Ética europeia sobre o uso de inteligência artificial em sistemas judiciais e em seu ambiente.** Foz do Iguaçu, 2019. Painel VI: Inteligência Artificial. Tradução livre de Teresa Germana Lopes de Azevedo. Disponível em <https://emeron.tjro.jus.br/images/biblioteca/revistas/Avulso/CartaEticaEuropeia.pdf> Acesso em 03 de Jun 2021.

CAMARGO, Gustavo Xavier de. Decisões judiciais computacionalmente fundamentadas: uma abordagem a partir do conceito de *explainable artificial intelligence*. **Revista Democracia Digital e Governo Eletrônico**, Florianópolis, v. 1, n. 18, p. 167-177, 2019.

CASTAGNA, Fabiano Pires. **Capacidade contributiva e igualdade tributária no imposto sobre a renda da pessoa física: os desafios da concretização sob a perspectiva do valor-princípio da fraternidade**. Tese de doutorado. UFSC, CCJ, PPGD, 2019.

COELHO, Eleonora. Desenvolvimento da cultura dos métodos adequados de solução de conflitos: uma urgência para o Brasil. **Arbitragem e mediação – a reforma da legislação brasileira**. Coordenadores: Caio Cesar Vieira Rocha e Luis Felipe Salomão. São Paulo: Atlas: 2015.

CONEIN, B.; COURTINE, J.-J.; GADET, F.; MARANDIN, J.-M.; PÊCHEUX, M. (orgs.). **Materialidades discursivas.** Campinas, SP: Editora da Unicamp, 2016.

CONSELHO NACIONAL DE JUSTIÇA (CNJ), **Resolução n. 332, publicada em 25 de agosto de 2020**. Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências. Disponível em <https://atos.cnj.jus.br/atos/detalhar/3429> Acesso em 03 jun 2021.

COURTINE, Jean-Jacques. Definição de orientações teóricas e construção de procedimento em Análise do Discurso. Tradução de Flávia Clemente de Souza e Márcio Lázaro Almeida da Silva. **Policromias – Revista de Estudos do Discurso, Imagem e Som**. Vol. 1. N. 1. Junho/2016. Disponível em < https://revistas.ufrj.br/index.php/policromias/article/view/4090 > Acesso em 31 jul 2020.

DIAS, Cristiane. **A materialidade digital da mobilidade urbana: espaço, tecnologia e discurso.** Línguas e instrumentos linguísticos, n. 37, jan-jun 20106. Disponível em < http://www.revistalinguas.com/edicao37/artigo7.pdf > Acesso em 28 Jun 2020.

DIAS, Cristiane e COUTO, Olivia Ferreira do. **As redes sociais na divulgação e formação do sujeito do conhecimento: compartilhamento e produção através da circulação de ideias.** *Ling. (dis)curso* [online]. 2011, vol.11, n.3, pp.631-648. Disponível em https://www.scielo.br/scielo.php?script=sci_arttext&pid=S151876322011000300009&lng=en & nrm=iso&tlng=pt Acesso em 15 Fev 2021.

DIMOULIS, Dimitri. **Manual de introdução ao estudo do direito**. 9 ed. São Paulo: Thomson Reuters Brasil, 2020.

EX_MAQUINA: INSTINTO ARTIFICIAL. Direção: Alex Garland. Produção de Scott Rudin, Eli Bush e Tessa Ross. Reino Unido: Universal Pictures, 2015. Telecineplay.

EXPRESSIONISMO ABSTRATO: O QUE É, CARCTERÍSTICAS, OBRAS E ARTISTAS. laart.art.br 2020. Disponível em < https://laart.art.br/blog/expressionismo-abstrato/> Acesso em 23 Jul 2020.

FARMAKIS, Christian. **Artificial intelligence is transforming the legal industry**. **Law,** Nova York, 31 jan 2020.

FERNANDES, Carolina; VINHAS Luciana Iost. **Da maquinaria ao dispositivo teórico-analítico: a problemática dos procedimentos metodológicos da Análise do Discurso.** Linguagem em (Dis)curso – LemD, Tubarão, SC, v. 19, n. 1, p. 133-151, jan./abr. 2019.

FLORES, Giovanna G. Benedetto. **Os sentidos de nação, liberdade e independência na imprensa brasileira (1821-1822) e a fundação do discurso jornalístico brasileiro.** Porto Alegre: EDIPUCRS; Palhoça: UNISUL, 2014.

FRIEDE, Reis. **Ciência do direito, norma, interpretação e hermenêutica jurídica**. 9ª ed. São Paulo: Manole, 2015.

GALLO, Solange Maria Leda. **Autoria: questão enunciativa ou discursiva?** In: Revista Linguagem em (Dis)curso, volume 1, número 2, jan./jun. 2001. Disponível em < http://www.portaldeperiodicos.unisul.br/index.php/Linguagem_Discurso/article/view/172 > Acesso em 28 Jul 2020

GALLO, Solange Maria Leda. **Novas fronteiras para a autoria**. Organon – Revista do Instituto de Letras da UFRGS, Porto Alegre, n. 53, julho-dezembro 2012, p. 53-64. Disponível em < https://seer.ufrgs.br/organon/article/view/35724 Acesso em 28 Jul 2020

GALLO, Solange Maria Leda. **Processo de legitimação no discurso de escritoralidade**. In: GRIGOLETTO, Evadra; STOCKMNAS DE NARDI, Fabiele (orpgs.). A Análise do Discurso e sua História. Campinas: Ed. Pontes, 2016.

GALLO, Solange Maria Leda. **Discurso e novas tecnologias da informação.** In: NAVARRO, P; BARONAS, R (Org.) – Sujeito, Texto e Imagem em Discurso. Ed. Pontes: Campinas, 2018.

GALLO, Solange Maria Leda; SILVEIRA, Juliana da. **Forma-discurso de oralidade: processos de normatização e legitimação.** In. FLORES, G; GALLO, S.; LAGAZZI, S.; NECKEL, N.; PFEIFFER, C.; ZOPPI-FONTANA (Org.). Análise de Discurso em rede: cultura e mídia Volume 3, 2017.

GILLET, Sérgio Augusto da Costa; PORTELA, Vinícius José Rockenbach. **Inteligência artificial e motivação das decisões judiciais: limites e desafios para a atividade cognoscitiva do juízo.** Processo e tecnologia. HOLZ, Jonathan Carvalho; MACEDO, Ekaibe Harzheim; GILLET Sérgio Augusto da Costa (Org.). Porto Alegre: Fi, 2018.

GUILHAUMOU, Jacques. **Linguística e história: percursos analíticos de acontecimentos discursivos.** Coordenação e organização da tradução Roberto Leiser Baronas e Fábio César Montanheiro. São Carlos: Pedro & João Editores, 2009.

GRILLO, Brenno. Excesso de plataformas de processo eletrônico atrapalha advogados. **Consultor Jurídico**. 2017. Disponível em https://www.conjur.com.br/2017-out-03/excesso-sistemas-processo-eletronico- atrapalham-advogados?imprimir=1. Acesso em 13/02/2021.

INDURSKY, Freda. **A fala dos quartéis e as outras vozes**. 2 ed. Campinas: Unicamp, 2013.

INDURSKY, Freda. A memória na cena do discurso. In: INDURSKY et al (org) **Memória e história na/da Análise do Discurso**. Campinas: Mercado de Letras, 2011.

Inteligência Artificial – Projeto Victor STF – Conversa com Dr. Fabiano Hartmann. [Locução de] Fabiano Hartmann Peixoto. Brasília: IAJUS-TEAM, 05 de outubro de 2019. *Podcast*. Disponível em < https://open.spotify.com/episode/1CPnWpFs8NJVlQLpwqlrxW?si=8Btsq8H0RSuDz0P_v16lMA > Acesso em 18/01/2021.

JUSTIÇA EXPRESS. **Ajudando pessoas e empresas a resolverem demandas jurídicas. Foi lesado? 2020.** Disponível em < https://hmg.justicaexpress.com> Acesso em 15 Jul 2020.

KELSEN, Hans. **Teoria pura do direito**. 7. ed. São Paulo: Martins Fontes, 2006.

KNEBEL, Patrícia. **Lei inédita de Inteligência Artificial entra em consulta pública**. Jornal do comércio, 2020. Disponível em < https://www.jornaldocomercio.com/_conteudo/colunas/mercado_digital/2020/07/747882-lei-inedita-de-inteligencia-artificial-entra-em-consulta-publica.html > Acesso em 23 Jul 2020.

LAGO, Laurenio. **Supremo Tribunal de Justiça e Supremo Tribunal Federal:** dados biográficos 1828-2001. 3. ed. Brasília: Supremo Tribunal Federal, 2001. p. 367-370.

LAGE, Fernanda de Carvalho. **Manual de inteligência artificial no direito brasileiro**. Salvador: Juspodivm, 2021.

MAIA FILHO, M. S.; JUNQUILHO, T. A. Projeto Victor: perspectivas de aplicação da inteligência artificial ao direito. **Revista de Direitos e Garantias Fundamentais**, v. 19, n. 3, p. 218-237**, 29 dez. 2018**. Disponível em https://sisbib.emnuvens.com.br/direitosegarantias/article/view/1587/pdf Acesso em 04/01/2021.

MITTMANN, Solange. **Discurso e texto: na pista de uma metodologia de análise**. In: INDURSKY, Freda; LEANDRO FERREIRA, Maria Cristina (Orgs.). *Análise do Discurso no Brasil*: mapeando conceitos, confrontando limites. São Carlos: Claraluz, 2007.

NUNES, José Horta. **A divulgação cientifica no jornal: ciência e cotidiano**. Produção e circulação do conhecimento. GUIMARÃES, Eduardo (Org.). Campinas: Pontes, 2003.

GILLET, Sérgio Augusto da Costa; PORTELA, Vinícius José Rockenbach. **Inteligência artificial e motivação das decisões judiciais: limites e desafios para a atividade cognoscitiva do juízo.** Processo e tecnologia. HOLZ, Jonathan Carvalho; MACEDO, Ekaibe Harzheim; GILLET Sérgio Augusto da Costa (Org.). Porto Alegre: Fi, 2018.

GONÇALVES, Marcus Vinicius Rios. **Direito processual civil esquematizado**. Coordenador Pedro Lenza. 7 ed. São Paulo: Saraiva, 2016.

HAROCHE, Claudine. **Fazer dizer, querer dizer.** Tradução de Eni Pulcinelli Orlandi, com a colaboração de Freda Insdursky e Marise Manoel. São Paulo: Hucitec, 1992.

MARTINS, Humberto. **Inteligência artificial aplicada à gestão de conflitos no âmbito do Poder Judiciário**. 1º Fórum sobre Direito e Tecnologia. FGV Conhecimento, Centro de Inovação, Administração e Pesquisa do Judiciário. Org. Luis Felipe Salomão. Painel 1. Disponível em < https://ciapj.fgv.br/sites/ciapj.fgv.br/files/anais_i_forum_ia.pdf> Acesso em 03 jun 2021.

OLIVEIRA, Rafael Tomaz. **Hermenêutica e jurisprudência no Novo Código de Processo Civil: a abertura de novos horizontes interpretativos no marco da integridade do direito**. *In:* STRECK, Lênio, ALVIM; Eduardo Alvim; e, LEITE, George Salomão (coords.). Hermenêutica e jurisprudência no Código de Processo Civil: coerência e integridade. 2. ed. São Paulo: Saraiva Educação, 2018.

ORLANDI, Eni Puccinelli. **Segmentar ou recortar?** Linguística: questões e controvérsias. Curso de Letras, Centro de Ciências Humanas e Letras das Faculdades Integradas de Uberaba, 1984 [Série Estudos].

ORLANDI, Eni Puccinelli. **Discurso e texto: formação e circulação dos sentidos.** Campinas: Pontes, 2012.

ORLANDI, Eni Puccinelli. **Análise de discurso. Princípios e procedimentos.** 12 ed. Campinas: Pontes, 2015.

ORLANDI, Eni Puccinelli. Nota introdutória à tradução brasileira. In: CONEIN, Bernard et atl. **Materialidades discursivas.** Campinas: Unicamp, 2016.

ORLANDI, Eni Puccinelli. A materialidade do gesto de interpretação e o discurso eletrônico. In. DIAS, Cristiane. **Formas de mobilidade no espaço e-urbano: sentido e materialidade digital** [online]. Série e-urbano. Vol. 2, 2013, Disponível em < https://www.labeurb.unicamp.br/livroEurbano/volumeII/arquivos/pdf/eurbanoVol2_EniOrlandi.pdf > Acesso em 27 Jul 2020.

ORLANDI, Eni Puccinelli. **Divulgação científica e efeito leitor: uma política social urbana**. Discurso e texto: formulação e circulação dos sentidos. 4 ed. Campinas: Pontes, 2012.

ORLANDI, Eni Puccinelli. **Interpretação: autoria, leitura e efeitos do trabalho simbólico**. 5 ed. Campinas: Pontes, 2020.

ORLANDI, Eni Puccinelli. **Discurso e leitura**. 9 ed. Campinas: Editora da Unicamp, 2012.

MAINGUENEAU, Dominique. **Novas tendências em análise do discurso.** Tradução Freda Indursky; revisão dos originais da tradução Solange Maria Leda Gallo, Maria da Glória de Deus Vieira de Moraes. 3 ed. Campinas: Pontes, 1997.

MASSON, Nathalia. **Manual de direito constitucional.** 9 ed. rev, ampl. e atual. Salvador: JusPODIVM, 2021.

MONTORO, André Franco. **Introdução à ciência do direito**. 27 ed. rev. e atual. São Paulo: RT, 2008.

OCDE, Organização para Cooperação e Desenvolvimento Econômico. **Recomendação do Conselho sobre Inteligência Artificial** (OCDE/LEGAL/0449). Paris, 21 de maio de 2019. Disponível em <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449#:~:text=The%20OECD%20Council%20adopted%20the,on%2022%2D23%20May%202019.&text=The%20OECD%20Recommendation%20on%20AI,governments%20in%20their%20implementation%20efforts.> Acesso em 03 de Jun 2021.

PÊCHEUX, Michel; LEON, Jacqueline; BONNAFOUS, Simone; MARANDIN, Jean-Marie. Apresentação da análise automática do discurso (1982). In. GADET, Françoise; HAK, Tony; tradução MARIANI, Bethania S. [et al.] **Por uma análise automática do discurso: uma introdução à obra de Michel Pêcheux.** Campinas: Unicamp, 2014.

PÊCHEUX, Michel; FUCHS, Catherine. A propósito da análise automática do discurso: atualização e perspectivas. In. GADET, Françoise; HAK, Tony; tradução MARIANI, Bethania S. [et al.] **Por uma análise automática do discurso: uma introdução à obra de Michel Pêcheux.** Campinas: Unicamp, 2014.

PÊCHEUX, Michel. **O Discurso: Estrutura ou Acontecimento.** Tradução de Eni Orlandi. 6ª ed., Campinas: Pontes, 2012.

PÊCHEUX, Michel. **Semântica e discurso – uma critica à afirmação do óbvio.** 5 ed. Campinas: Unicamp, 2014.

PÊCHEUX, Michel. Análise Automática do Discurso (1969). Trad. Eni Orlandi. In: GADET, Françoise & HAK, Tony. (Orgs.). **Por uma Análise Automática do Discurso: uma introdução à obra de Michel Pêcheux**. 5ª edição. Campinas-SP: Ed. da Unicamp, 2014.

PEQUENO, Vitor. **Nos subsolos de uma rede: sobre o ideológico no âmago do técnico.** Dissertação (Mestrado) - Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, SP, 2015.

PEQUENO, Vitor. **Tecnologia e esquecimento: uma crítica a representações universais de linguagem.** 2019. 224 f. Tese (Doutorado) - Curso de Linguística, Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, SP, 2019.

PEITOXO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. **Inteligência artificial e direito**. Curitiba: Alteridade, 2019.

PEIXOTO, Fabiano Hartmann. **Projeto Victor: relato do desenvolvimento da Inteligência Artificial na Repercussão Geral do Supremo Tribunal Federal.** Revista Brasileira de Inteligência Artificial e Direito. ISSN 2675 - 3156. v. 1, n. 1, Jan – Abr., 2020, p. 1 - 20.

PIRES, Carolina. **Tese estuda projeto pioneiro da UnB de inteligência artificial para o Poder Judiciário**. UnBCIÊNCIA - Humanidades, **2020.** Disponível em http://www.unbciencia.unb.br/humanidades/57-direito/661-tese-estuda-projeto-pioneiro-de-inteligencia-artificial-para-o-poder-judiciario Acesso em 10/01/2021

PÊCHEUX, Michel. Ler o arquivo hoje. **Gestos de Leitura: da história no discurso**. Eni P. Orlandi (org.) [et al.] 4ª ed. Campinas: UNICAMP, 2014, p. 63.

PIOVEZANI, Carlos; SARGENTINI, Vanice (Org.). **Legados de Michel Pêcheux: inéditos em análise do discurso.** São Paulo: Contexto, 2011.

REALE, Miguel. **Teoria tridimensional do direito.** 5. ed. rev. e reest. São Paulo: Saraiva, 1994.

ROQUE, André Vasconcelos; SANTOS, Lucas Braz Rodrigues dos. **INTELIGÊNCIA ARTIFICIAL NA TOMADA DE DECIÕES JUDICIAIS: TRÊS PREMISSAS BÁSICAS.** Revista Eletrônica de Direito Processual – REDP. Rio de Janeiro. Ano 15. Volume 22. Número 1. Janeiro a Abril de 2021. Disponível em https://www.e-publicacoes.uerj.br/index.php/redp/article/view/53537/36309 Acesso em 10/01/2021

SALOMÃO, Luis Felipe (Org). **Inteligência artificial aplicada à gestão dos conflitos no âmbito do Poder Judiciário.** 1º Fórum sobre direito e tecnologia. FGV Conhecimento – Centro

de Inovação, Administração e Pesquisa do Judiciário. Publicado em 02/12/2020. Disponível em https://ciapj.fgv.br/sites/ciapj.fgv.br/files/anais_i_forum_ia.pdf Acesso em 14 Fev 2021.

SAUSSURE, Ferdinand de. **Curso de linguística geral.** 27. ed. São Paulo: Cultrix, 2006.

SELEME, Mariana Pigatto e SOUZA, Marina Haline. **CNJ publica resolução sobre produção e uso de inteligência artificial no Poder Judiciário**. Migalhas de peso, publicada em 10 set 2020. Disponível em <https://www.migalhas.com.br/depeso/333093/cnj-publica-resolucao-sobre-producao-e-uso-de-inteligencia-artificial-no-poder-judiciario> Acesso em 03 jun 2021.

SILVA Sobrinho, Helson Flávio da. **AAD-69: uma referência incontornável**. Línguas e instrumentos linguísticos. N. 44, jul-dez 2019. Disponível em <http://www.revistalinguas.com/edicao44/resenha.pdf> Acesso em: 09 abril 2021.

SILVEIRA, Juliana da. ***Hashtags* e *trending topics*: a luta pelo(s) sentido(s) nos espaços enunciativos informatizados: a luta pelo(s) sentido(s) nos espaços enunciativos informatizados.** Interletras, Grande Dourados, v. 31, n. 8, p. 1-18, 31 abr. 2020. Disponível em <https://tinyurl.com/yabu9urb.> Acesso em: 19 junho 2020.

SOARES, Ricardo Maurício Freire. **Hermenêutica jurídica.** São Paulo: Saraiva 2012. Recurso online.

SOUZA, Lara Oliveira. **Princípio da transparência algorítmica e devido processo legal: um diálogo necessário para garantia do direito à explicação**. A natureza e o conceito do direito 3 [recurso eletrônico] / Organizador Adaylson Wagner Sousa de Vasconcelos. – Ponta Grossa, PR: Atena, 2020. Disponível em < https://www.atenaeditora.com.br/post-artigo/41489 > Acesso em: 17 abril 2021.

STF. **Inteligência artificial vai agilizar a tramitação de processos no STF.** Publicada em 30 de maio de 2018. Disponível em <http://www.stf.jus.br/portal/cms/verNoticiaDetalhe. asp?idConteudo=380038> Acesso em 03 de abr de 2021.

STF. **Ministra Cármen Lúcia anuncia início de funcionamento do Projeto Victor, de inteligência artificial**. Publicada em 30 de agosto de 2018. Disponível em

http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=388443 Acesso em 14 de Jul de 2020.

STF. **Projeto VICTOR do STF é apresentado em congresso internacional sobre tecnologia**. Publicada em 26 de setembro de 2018. Disponível em http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=388443 Acesso em 15 de Jul de 2020.

STF. **Presidente do Supremo apresenta ferramentas de inteligência artificial em Londres**. Publicada em 05 de setembro de 2019. Disponível em http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=422699 Acesso em 14 Fev de 2021.

STF. **Ministros – Victor Nunes Leal,** disponível em http://www.stf.jus.br/portal/ministro/verMinistro.asp?periodo=stf&id=108 Acesso em 12 Fev 2021.

UNESCO inicia consulta pública sobre padrões globais de ética na inteligência artificial. **Nações Unidas Brasil.** 2020. Disponível em < https://nacoesunidas.org/unesco-inicia-consulta-publica-sobre-padroes-globais-de-etica-na-inteligencia-artificial/ > Acesso em 1º Ago 2020.

VIEIRA, Leonardo Marques**. A problemática da inteligência artificial e dos vieses algorítmicos: caso compas**. Brazilian Technology Symposium. UNICAMP: 2019. Disponível em https://www.lcv.fee.unicamp.br/images/BTSym-19/Papers/090.pdf Acesso em 24.03.2021

ANEXOS

United Nations Educational, Scientific and Cultural Organization

World Commission on the Ethics of Scientific Knowledge and Technology (COMEST)

SHS/COMEST/EXTWG-ETHICS-AI/2019/1
Paris, 26 February 2019
Original: English

PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE

Building on the work of COMEST on Robotics Ethics (2017) and on the Ethical Implications of the Internet of Things (IoT), this preliminary study is prepared by a COMEST Extended Working Group on Ethics of Artificial Intelligence.

This document does not claim to be exhaustive and does not necessarily represent the views of the Member States of UNESCO.

PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE

TABLE OF CONTENT

PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE

INTRODUCTION

1. The world is facing a rapid rise of 'Artificial Intelligence' (AI). Advancements in this field are introducing machines with the capacity to learn and to perform cognitive tasks that used to be limited to human beings. This technological development is likely to have substantial societal and cultural implications. Since AI is a cognitive technology, its implications are intricately connected to the central domains of UNESCO: education, science, culture, and communication. Algorithms have come to play a crucial role in the selection of information and news that people read, the music that people listen to, and the decisions people make. AI systems are increasingly advising medical doctors, scientists, and judges. In scientific research, AI has come to play a role in analysing and interpreting data. Furthermore, the ongoing replacement of human work by intelligent technologies demands new forms of resilience and flexibility in human labour. Public thinkers like Stephen Hawking have even voiced the fear that AI could bring an existential threat to humankind, because of its potential to take control of many aspects of our daily lives and societal organization.

2. In the 1950s, the term 'artificial intelligence' was introduced for machines that can do more than routine tasks. As computing power increased, the term was applied to machines that have the ability to learn. While there is not one single definition of AI, it is commonly agreed upon that machines which are based on AI, or on 'cognitive computing', are potentially capable of imitating or even exceeding human cognitive capacities, including sensing, language interaction, reasoning and analysis, problem solving, and even creativity. Moreover, such 'intelligent machines' can demonstrate human-like learning capabilities with mechanisms of self-relation and self-correction, on the basis of algorithms that embody 'machine learning' or even 'deep learning', using 'neural networks' that mimic the functioning of the human brain.

3. Recently, large multinational tech companies in many regions of the world have started to invest massively in utilizing AI in their products. Computing power has become large enough to run highly complicated algorithms and to work with 'big data': huge sets of data that can be used for machine learning. These companies have access to almost unlimited computing power and also to data collected from billions of people to 'feed' AI systems as learning input. Moreover, via their products, AI is rapidly gaining influence in people's daily lives and in professional fields like healthcare, education, scientific research, communications, transportation, security, and art.

4. This profound influence of AI raises concerns that could affect the trust and confidence people have in these technologies. Such concerns range from the possibility of criminality, fraud and identity theft to harassment and sexual abuse; from hate speech and discrimination to the spreading of disinformation; and more generally from the transparency of algorithms to the possibilities of trusting AI systems. Since many of these problems cannot be addressed by regulation alone, UNESCO has been proposing multi-stakeholder governance as an optimum modality to involve the various actors in the formulation and implementation of norms, ethics and policy, as well as the empowerment of users.

5. Because of its profound social implications, many organizations and governments are concerned about the ethical implications of AI. The European Commission has formed a High Level Expert Group on AI comprising representatives from academia, civil society, industry, as well as a European AI Alliance, which is a forum engaged in broad and open

discussion on all aspects of AI development and its impacts. The European Group on Ethics in Science and New Technologies has issued a *Statement on AI, Robotics, and Autonomous Systems* (EGE, 2018). The European Commission has published a *Communication on AI for Europe* (EC, 2018) and the Council of Europe has produced various reports on AI and has formed a Committee of Experts to work on the *Human rights dimensions of automated data processing and different forms of artificial intelligence*. The IEEE organization has formed a Global Initiative on Ethics of Autonomous and Intelligent Systems. The OECD has initiated the 'Going Digital' project, which aims to help policymakers in all relevant policy areas better understand the digital revolution that is taking place across different sectors of the economy and society as a whole. The OECD has also created an expert group (AIGO) to provide guidance in scoping principles for artificial intelligence in society. ITU and the WHO have established a Focus Group on "Artificial intelligence for Health". Furthermore, many countries have initiated reflection on their ethical and political orientation towards AI, like the Villani report in France (Villani et al., 2018); the House of Lords report in the UK (House of Lords, 2017); and the report of the Executive Office of the President of the USA (2016).

6. UNESCO has a unique perspective to add to this debate. AI has implications for the central domains of UNESCO's work. Therefore, in addition to the many ethical guidelines and frameworks that are currently being developed by governments, companies, and societal organizations, UNESCO can bring a multidisciplinary, universal and holistic approach to the development of AI in the service of humanity, sustainable development, and peace.

7. In this regard, there are several existing frameworks and initiatives to build on. Firstly, there is the *human rights* framework, which formed the basis of the 2003 World Summit on the Information Society's (WSIS) Geneva Declaration of Principles, stating that "the use of ICTs and content creation should respect human rights and fundamental freedoms of others, including personal privacy, and the right to freedom of thought, conscience, and religion in conformity with relevant international instruments" (WSIS, 2003). WSIS (2005) proposes a multi-stakeholder approach that calls for an effective cooperation of all stakeholders, including Governments, the private sector, civil society, international organizations, and the technical and academic communities. In the WSIS follow-up process, UNESCO has adopted this multi-stakeholder approach and has taken responsibility for the implementation of the Action Lines on Access (C3), E-Learning (C7), Cultural diversity (C8), Media (C9), and Ethical dimension of the information society (C10).

8. Second, there is the framework of *Internet Universality* and the associated *R.O.A.M. principles* as approved by the 38th General Conference in 2015 (UNESCO, 2015b). These principles cover Human Rights, Openness, Accessibility and Multi-stakeholder participation, and have emerged from the UNESCO "Keystones" study for the 38th General Conference (UNESCO, 2015a). In the "Connecting the Dots" outcome document of this conference, UNESCO commits to promoting human rights-based ethical reflection, research and public dialogue on the implications of new and emerging technologies and their potential societal impacts. Moreover, the 18th Session of the Intergovernmental Council of UNESCO's Information For All Programme (IFAP) examined and approved the Code of Ethics for the Information Society, which was elaborated by the IFAP Working Group on Information Ethics.

9. To investigate the ethical implications of AI, this study will first explain what Artificial Intelligence is, how it works, and how it is different from other technologies. The second section will investigate the ethical aspects of AI, taking the UNESCO domains of education, science, culture, and communication as a starting point, as well as the global-

ethical dimensions of peace, cultural diversity, gender equality, and sustainability. This investigation should be seen as an exploration, not as a comprehensive analysis, ranging from cultural diversity to trust in science, from artistic creativity to critical thinking, and from AI-based decision-making to the role of AI in developing countries. The third section of this preliminary study will sketch out the central dimensions that proper ethical reflection on AI should have from the perspective of UNESCO.

I. WHAT IS AI?

I.1. Definition

10. The idea of 'artificial intelligence' (AI) – as the idea of 'artificially created' and 'intelligent' beings, machines or tools – is scattered throughout human history. Its various forms can be found in both Western and non-Western religions, mythologies, literature and philosophical traditions. As such, these instances testify to the perennial curiosity of humankind with such entities, and despite the expression of this curiosity through culturally diverse appearances, it is something shared or cross-cultural. Today, the fascination with AI – including its ethical dimensions – is amplified by its development and real-world applications.

11. Any examination of the ethical implications of AI need a clarification of its possible meanings. The term was coined in 1955, by John McCarthy, Marvin L. Minsky, Nathaniel Rochester and Claude E. Shannon. The 'study of artificial intelligence' was planned "to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it" (McCarthy et al., 2006 [1955], p.12). As the field developed and diversified in the decades to come, the number of meanings of 'AI' increased and there is no universally agreed upon definition today. Various definitions of AI are related to different disciplinary approaches such as computer science, electrical engineering, robotics, psychology or philosophy.

12. Despite the multitude and diversity of definitions of AI, there is certain consensus, at the most general level, that its two aspects can be distinguished: one usually labelled as 'theoretical' or 'scientific' and the other one as 'pragmatic' or 'technological'.

13. To talk about 'theoretical' or 'scientific' AI is about "using AI concepts and models to help answer questions about human beings and other living things" (Boden, 2016, p.2). 'Theoretical' or 'scientific' AI thus naturally interconnects with disciplines like philosophy, logic, linguistics, psychology and cognitive science. It deals with questions like: What is meant by 'intelligence' and how to distinguish 'natural' from 'artificial' intelligence? Is symbolic language necessary for thought processes? Is it possible to create 'strong AI' (*genuine* intelligence of the same kind and level of generality as human intelligence) as opposed to 'weak AI' (intelligence that only *mimics* human intelligence and is able to perform a limited number of narrowly defined tasks)? Although questions like these are theoretical or scientific, they involve a number of metaphysical or spiritual concerns (e.g. about human uniqueness or the freedom of will) which themselves have indirect, but nonetheless serious, ethical implications.

14. 'Pragmatic' or 'technological' AI is engineering-oriented. It draws on various branches of AI – textbook examples are natural language processing, knowledge representation, automated reasoning, machine learning, deep learning, computer vision and robotics (Russell and Norvig, 2016, p.2-3) – in order to create machines or programs capable of independently performing tasks that would otherwise require human

intelligence and agency. ‘Pragmatic’ or ‘technological’ AI became remarkably successful as they are combined with ICT (information and communications technology). AI innovations are used today in many areas of modern life, such as transport, medicine, communication, education, science, finance, law, military, marketing, customer services or entertainment. These innovations do raise direct ethical concerns, ranging from the disappearance of traditional jobs, over responsibility for possible physical or psychological harm to human beings, to general dehumanization of human relationships and society at large. At the moment, no AI system can be considered as a general-purpose intelligent agent that can perform well in a wide variety of environments, which is a proper ability of human intelligence.

15. One of the particularities of AI concerns its ‘unfamiliarity’ to us humans in a sense that the way its intelligence works seems strange and mysterious to us. The essence of this ‘unfamiliarity’ is what one might call ‘performance without awareness’. High-functioning AI such as AlphaGo or Watson can perform impressively without recognizing what it is doing. AlphaGo defeated a number of Go masters without even knowing that it was playing a human game called Go. Watson answered devilish questions so fast, which most humans have difficulties with even understanding in the given time. However, Watson is not ‘answering’ in the human sense; rather it ‘computes’ the probabilities of several candidate answers based on its automated analysis of an available database. AlphaGo and Watson perform brilliantly without being aware of what they are doing.

16. There are certainly important philosophical questions about whether the ‘play’ of AlphaGo and the ‘answer’ of Watson are ‘genuine’ or not. An ethically more crucial fact is however that we humans are not used to this kind of intelligence. Whenever we are confronted with impressive works of art, literature and science, we naturally consider the ‘conscious’ intelligence behind them. We recognize the unique character of Beethoven behind his 9th symphony, and the overwhelming searching mind behind Goedel’s incompleteness theorem. The simple fact that we should not apply this familiar rule of thumb in regard to brilliant performances when we interact with high-functioning AI poses serious social and ethical challenges. As we are used to interacting emotionally and socially with behaviourally intelligent agents, we naturally interact emotionally and socially with ‘high-functioning AI without awareness’, such as so-called ‘emotion’ or ‘social robots’- for example ‘smart home assistant’ (Alexa, Siri, Google assistant). At the current level of technological development, high-functioning AI without awareness cannot properly reciprocate complicated emotional and social expectations of human agents, while its external behaviour coupled with human imagination could generate an ‘unrealistic’ hope of genuine interactions with humans. It is important for us to remember that the seemingly ‘emotional’ mind of AI is much more of our imagination rather than of reality. There is general agreement that artificially intelligent systems do not have awareness in the experiential human sense, even if they can answer questions about the context of their actions. It is important not to equate experience with intelligence, even though some experts have suggested that recent developments in AI might also be a reason to re-examine the importance of this experience or awareness for being human. If experience is at the core of being human, ethical considerations must ensure that this is protected and enhanced through the use of AI rather than side-lined or disempowered. However, it may be that our experience with high-functioning AI without awareness can still influence our interactions with ordinary humans with awareness.

I.2. How does AI work?

17. To be able to perform the tasks of a human mind, an AI machine needs to be able

to sense the environment and to collect data dynamically, to process it promptly and to respond – based on its past 'experience', its pre-set principles for decision-making and its anticipation about the future. However, the technology behind AI is a standard ICT: it is based on collecting/acquiring data, storing, processing and communicating it. The unique features of cognitive machines come from quantities, which are transformed into qualities. AI technology is based on the following components:

a. *Dynamic data*. The system needs to be exposed to changing environments and to all relevant data acquired by various sensors, to classify and to store it, and to be able to process it promptly.
b. *Prompt processing*. Cognitive machines must react promptly. AI therefore needs to have reliable, fast and strong computing and communication resources.
c. *Decision-making principles*. AI decision-making is based on machine learning algorithms. Therefore, its response to a specific task depends on its 'experience' – that is, on the data it has been exposed to. The algorithms behind the decisions made by cognitive machines are based on some general principles the algorithm obeys and tries to optimize, given the data it is provided with.

The present ability to efficiently integrate dynamic data acquisition and machine-learning algorithms for prompt decision-making enables the creation of 'cognitive machines'.

I.3. How is AI different from other technologies?

18. Most 20th century technologies are model-driven. That is, scientists study nature and suggest a scientific model to describe it, and technology is advanced based on such models. For example, understanding the propagation of electromagnetic waves is the basis for the technology of wireless telecommunication. Modelling of the human brain is, however, a task which still seems far from being at a stage where a cognitive machine can be model-based. Therefore, AI is built on a different approach: a data-driven approach.

19. The data-driven approach is at the core of *machine learning*, which is commonly based on 'artificial neural networks' (ANNs). ANNs are formed by a series of nodes conceptually similar to brain neurons interconnected through a series of layers. The nodes of the input layer receive information from the environment, where, at each node, a non-linear transformation is applied. Such systems 'learn' to perform tasks by considering examples (labelled data), generally without being programmed with any task-specific rules or models. Deep learning, to conclude, is based on ANNs of several layers, which enables the machine to recognize complex concepts such as human faces, human bodies, speech understanding and all types of images classification.

20. The key issue in the ability of AI to show human-like capabilities is its scalability. The performance of AI machines depends on the data to which they are exposed, and for best performance, access to relevant data should be borderless. There may be technical limitations to the access to data, but the way data are selected and classified is also a socio-cultural issue (Crawford, 2017). Classification is culture-specific and a product of history, and may create bias in the decisions made by the algorithm. If the same machine is exposed to diverse sets of data, its bias can be reduced but not completely suppressed (Executive Office of the President, 2016). It is important to point out that in order to comply with what is mandated in Article 27 of the Universal Declaration of Human Rights – stating that every human being is entitled to the benefits of scientific progress – and to ensure

diversity in data sets available for AI, it is relevant to promote the capacity building of states, both in terms of human skills and infrastructure.

21. AI technology has matured under the drive of multinational companies that are not confined to local and national constraints. Moreover, to ensure prompt processing and reliability of the systems, the actual location of computing processes is distributed and the location of an AI machine is not defined by the place where it operates. Practically, AI is based on cloud technology, where the location of the storage and processing units can be anywhere. AI technology is characterized by the following:

a. While many of its applications are in the public sphere, AI technology is developed and led by multinational companies, most of them operating in the private sector and less obligated to the public good.
b. AI is not confined to a tangible location. This poses a challenge as regards how to regulate AI technology nationally and internationally.
c. The technology is based on accessibility to personal as well as public data.
d. AI technologies are not neutral, but inherently biased due to the data on which they are trained, and the choices made while training with the data.
e. AI and cognitive machine decisions cannot be fully predictable or explainable. Rather than operating mechanistically or deterministically, AI software learns from dynamic data as it develops and incorporates real-world experience into its decision-making.

II. ETHICAL CONSIDERATIONS

22. Artificial Intelligence has substantial societal and cultural implications. As many information technologies do, AI raises issues of freedom of expression, privacy and surveillance, ownership of data, bias and discrimination, manipulation of information and trust, power relations, and environmental impact in relation to its energy consumption. Moreover, AI brings specifically new challenges that are related to its interaction with human cognitive capacities. AI-based systems have implications for human understanding and expertise. Algorithms of social media and news sites can help to spread disinformation and have implications for the perceived meaning of ‘facts’ and ‘truth’, as well as for political interaction and engagement. Machine learning can embed and exacerbate bias, potentially resulting in inequality, exclusion and a threat to cultural diversity. The scale and the power generated by AI technology accentuates the asymmetry between individuals, groups and nations, including the so-called ‘digital divide’ within and between nations. This divide may be exacerbated due to lack of access to fundamental elements such as algorithms for learning and classification, data to train and to evaluate the algorithms, human resources to code, set up the software, and prepare the data, as well as computational resources for storage and processing of data.

23. As a result, Artificial Intelligence requires careful analysis. From UNESCO’s perspective, the most central ethical issues regarding Artificial Intelligence concern its implications for culture and cultural diversity, education, scientific knowledge, and communication and information. In addition to this, given UNESCO’s global orientation, the global-ethical themes of peace, sustainability, gender equality, and the specific challenges for Africa also deserve separate attention.

II.1. Education

24. Artificial Intelligence challenges the role of education in societies in many respects. Firstly, AI requires a rethinking of the societal role of education. The labour displacement caused by some forms of AI requires, among other measures, the retraining of employees, and a new approach to formulate the final qualifications of educational programmes. Moreover, in a world of AI, education should empower citizens to develop new forms of critical thinking, including 'algorithm awareness' and the ability to reflect on the impact of AI on information, knowledge, and decision-making. A second field of ethical questions regarding AI and education concerns its role in the educational process itself, as an element of digital learning environments, educational robotics, and systems for 'learning analytics', all of which require responsible development and implementation. Finally, engineers and software developers should be appropriately trained to ensure responsible design and implementation of AI.

II.1.1. The societal role of education

25. One of the main societal concerns regarding AI is labour displacement. The speed of change that AI is bringing presents unprecedented challenges (Illanes et al., 2018). It will involve, in the near future, the need to retrain large numbers of workers, and will have deep implications for the career paths students will need to follow. According to a McKinsey panel survey of 2017, "executives increasingly see investing in retraining and "upskilling" existing workers as an urgent business priority" (Illanes et al., 2018).

26. AI, therefore, will urge societies to rethink education and its social roles. Traditional formal education provided by universities might no longer be enough in the rise of digitized economies and AI applications. Until now, the standard education model has typically been to provide 'core knowledge' (Oppenheimer, 2018) and has focused on formal literacies like reading, writing and mathematics. In the 21st century, information and knowledge are omnipresent, demanding not only 'data literacy' that allows students to read, analyse and efficiently manage this information but also 'AI literacy' to enable critical reflection on how intelligent computer systems have been involved in the recognition of information needs, selection, interpretation, storage and representation of data.

27. Moreover, in a continuously developing labour market, the educational system can no longer aim to educate people for one specific profession. Education should enable people to be versatile and resilient, prepared for a world in which technologies create a dynamic labour market, and in which employees need to re-school themselves on a regular basis. Current ideas about 'lifelong learning' might need to be up-scaled into a model of continuous education, including the development of other types of degrees and diplomas.

II.1.2. AI in teaching and learning

28. Open educational resources (OER) have been an important addition to the learning landscape with the free availability of high quality lectures and other teaching resources through the internet. The potential of OERs to impact the education of people from across the world is unparalleled, but has yet to be fully realised as the limited completion rates for massive open online courses (MOOCs) demonstrates. The wide variety and depth of resources available has given rise to two problems. Firstly, the problem of finding the right resource for either an individual learner or a teacher wishing to reuse a resource in their own teaching materials. This has led to the second problem of reducing diversity through some resources becoming very popular at the expense of other potentially more relevant but less accessible content.

29. An example here is the Horizon 2020 project “X5GON” (Cross Modal, Cross Cultural, Cross Lingual, Cross Domain, and Cross Site Global OER Network: https://www.x5gon.org/). This project, funded by the European Union, is developing Artificial Intelligence methods to enable both learners and teachers to identify resources that match their learning goals, taking into account the specifics of their situation. For example, a teacher in Africa might be directed to lectures that present a topic based on local and indigenous knowledge that is appropriate for the particular cultural and local context, but equally would enable a learner from elsewhere interested in understanding specific African challenges to find relevant African content potentially translated from a local language.

30. In this way, AI can potentially address both of the above-identified problems. The first problem is tackled through assisting in the identification of the resources that are better matched to the learner’s or teacher’s needs through modelling their interests and goals, while at the same time exploiting an enriched representation of the huge repositories of OERs available throughout the world. By tuning recommendations to the individual learner or teacher, it further addresses the second problem, as the recommendations will no longer default to the most popular resource on a particular topic. There is the further potential to link learners from different cultures to enhance cross-cultural sharing of ideas and hence supporting mutual understanding and respect.

II.1.3. Educating AI engineers

31. The development of future technologies is in the hands of technical experts. Traditionally, engineers are educated to develop products to optimize performance using minimum resources (power, spectrum, space, weight etc.), under given external constrains. Over the past decades, the ethics of technology has developed various methods to bring ethical reflection, responsibility and reasoning to the design process. In the context of AI, the term ‘ethically aligned design’ (EAD) has been developed to indicate design processes that explicitly include human values (IEEE, 2018).

32. It is most important to apply Ethically Aligned Design in AI and other autonomous, intelligent systems (AIS) because this makes it possible to address ethical issues at a moment when the technology can still be adapted. A good example is ‘privacy by design’. Privacy can be violated less if not all data is stored but only that which is required for a specific task. An example of this is crowd counting, i.e. counting people in a crowd based on photos. In this case, if the photo is pre-processed to extract only the contours (edges) of the figures, people will remain unrecognizable and the counting algorithm will perform well without violating privacy. Similarly, AI developers can consider other ethical issues such as the prevention of algorithmic bias and traceability, minimizing the ability to misuse the technology, and explainability of algorithmic decisions.

33. Global engineering education today is largely focused on scientific and technological courses that are not intrinsically related to the analysis of human values overtly designed to positively increase human and environmental wellbeing. It is most important to change this and to educate future engineers and computer scientists for ethically aligned design of AI systems. This requires an explicit awareness of the potential societal and ethical implications and consequences of the technology-in-design, and of its potential misuse. The IEEE (a global organization of more than 400,000 electrical engineers) already promotes this issue via its global initiative on the ethics of autonomous and intelligent systems (https://ethicsinaction.ieee.org/). Addressing this issue is also a matter of ensuring active efforts for gender inclusion as well as social and cultural diversity of engineers, and for a holistic application of societal and ethical implications of AI system

design. Occasions for dialogue between engineers and the public should be encouraged in order to facilitate communication on the needs and visions of society, and on how engineers really work and conduct research in their everyday activities.

II.2. AI and Scientific Knowledge

34. In the field of scientific practice, AI is likely to have profound implications. In the natural and social sciences as well as in the life sciences and environmental sciences, it challenges our concepts of scientific understanding and explanation in a fundamental way. This also has implications on how we apply scientific knowledge in social contexts.

II.2.1. AI and scientific explanation

35. Because of the increasingly powerful forms of machine learning and deep learning, AI challenges existing conceptions of satisfactory scientific explanation as well as what we can naturally expect from predictably successful scientific theories. In the conventional view of science, the so-called deductive-nomological model, a proper scientific explanation is able to make correct predictions of specific phenomena based on scientific laws, theories and observations. For instance, we can legitimately say that we explain how the moon is moving around the earth in terms of Newtonian mechanics only when we are able to employ Newtonian mechanics in a deductive way to predict the lunar orbit. Such predictions are typically based on causal understanding, or on a unifying understanding of seemingly disparate phenomena.

36. In contrast to this, AI can reliably produce impressively accurate predictions based on data sets without giving us any causal or unifying explanation of its predictions. Its algorithms do not work with the same semantic concepts that humans employ to achieve scientific understanding of a phenomenon. This gap between successful predictions on the one hand and satisfactory scientific understanding on the other is likely to play a key role in scientific practice, as well as in decision-making based on AI.

37. This might have implications for trust in science, which is typically based on the scientific method that explains different phenomena in a systematic and transparent way, making its predictions rational and evidence-based. The apparent success of machine learning algorithms to deliver comparable results without such a scientifically justified model could have implications for the public perception and evaluation of science and scientific research.

38. Moreover, research shows that the quality of machine learning depends heavily on the available data used to train the algorithms. But since most AI applications are developed by private companies, there is not always enough transparency about these data, in contrast to the traditional scientific method that warrants the validity of results by requiring replicability, i.e. the possibility to reproduce them by repeating the same experiments.

II.2.2. AI, life sciences and health

39. Within the life sciences and medicine in particular, the development of AI technologies has significantly transformed the health care and bioethics landscape over the years. They can bring positive effects, like more precision in robotic surgery, and better care for autistic children, but at the same time, they raise ethical concerns, such as the cost they bring within the context of scarcity of resources in the health care system and the transparency they should bring in order to respect the autonomy of patients.

40. From an individual perspective, AI is bringing a new way of dealing with health and medical issues for the lay public. The use of internet sites and the multiplication of mobile phone software applications for self-diagnosis have given people the opportunity to generate health diagnoses without the participation of a health professional. This might have implications for medical authority and for the acceptance of self-medication, including the dangers it entails. It also changes the doctor-patient relationship, and calls for some kind of regulation without hindering innovation and autonomy.

41. AI technologies might free up time for health providers to dedicate to their patients, for instance by facilitating data entry and deskwork, but at the same time, they might replace the holistic and human elements of care. The well-known technology Watson for Oncology by IBM is a breakthrough in cancer treatment, but also raises important questions regarding the character and expectations of medical expertise and education, and the responsibilities of doctors working with the system. Similar concerns are raised with the development of chatbots for people seeking psychological help and counselling, apps for early detection of episodes of psychiatric diseases, or AI systems for producing psychiatric diagnoses on the basis of information collected from people's activity on social media and the Internet – which obviously also has important implications for privacy. In addition, in the case of the elderly, AI-based technologies such as assistive social robots are being introduced which can be useful on medical grounds for patients with dementia for example, but also raise concerns about reduced human care and the resulting social isolation.

42. AI also brings a new dimension to the ongoing discussion about 'human enhancement' versus 'therapy'. There are initiatives to integrate AI with the human brain using a 'neural interface': a mesh growing with the brain, which would serve as a seamless brain-computer interface, circulating through the host's veins and arteries (Hinchliffe, 2018). This technological development has important implications for the question of what it means to be human, and what 'normal' human functioning is.

II.2.3. AI and environmental science

43. AI has the potential to be beneficial to environmental science through a number of different applications. It can be used to process and interpret data within ecology, systems biology, bioinformatics, space and climate research, thus enhancing scientific understanding of processes and mechanisms. Improved recycling, environmental monitoring and remediation, and more efficient energy consumption can have direct environmental benefits. AI in agriculture and farming can lead to improved crop production (e.g., automated fertilization and irrigation) and animal welfare, and reduced risks from disease, pests, or weather threats. On the other hand, AI could lead to changes in human perceptions of nature, either positively by enhancing human awareness of beauty or independency, or negatively through increased 'instrumentalization' of nature or separation between humans and animals or the environment.

44. For all applications, the potential benefits need to be balanced against the environmental impact of the entire AI and IT production cycle. This includes mining for rare-earth elements and other raw materials, the energy needed to produce and power the machines, and the waste generated during production and at the end of life cycles. Increased AI is likely to add to the growing concerns about the increasing volumes of e-waste and the pressure on rare-earth elements generated by the computing industry. In addition to the environmental and health impacts, e-waste has important socio-political implications, especially related to the export to developing countries and vulnerable populations (Heacock et al., 2015).

45. Disaster risk management is an area where AI can aid in the prediction and response to environmental hazards such as tsunamis, earthquakes, tornadoes and hurricanes. A concrete example is the UNESCO G-WADI Geoserver application (Water and Development Information), which is being used to inform emergency planning and management of hydrological risks, such as floods, droughts and extreme weather events. Its support system PERSIANN (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks) is a satellite-based precipitation retrieval algorithm, providing near-real time information. The PERSIANN Cloud Classification System CCS algorithm (accessible at: http://hydis.eng.uci.edu/) has been optimized for observing extreme precipitation, particularly at very high spatial resolution and is being widely used globally to track storms. It also offers an iRain mobile application (http://en.unesco.org/news/irain-new-mobile-app-promote-citizen-science-and-support-water-management) where crowdsourcing gives opportunities for engaging citizen scientists in data collection.

46. Interestingly, even private companies have recently contributed to disaster management. One such example is Google's AI-enabled flood forecasting project (https://www.blog.google/products/search/helping-keep-people-safe-ai-enabled-flood-forecasting/). In this regard, the development of AI technologies that could bring potential benefit for disaster management should be encouraged.

II.2.4. AI and social sciences

47. Broadly speaking, social science research aims at finding out the causal structure of personal and social interactions. As most social phenomena are multiply influenced by a number of causal factors, social scientists typically rely on statistical analysis of the relevant empirical data to determine prominent causal factors and the strength of their effects. While doing so, it is crucial to distinguish mere statistical correlations from genuine causal connections. Certainly AI has clear potential to help social scientists navigate huge data sets to come up with plausible causal mechanisms as well as verify the validity of the proposed ones. On the other hand, AI can 'overfit' the data, and put forward 'pseudo' causal relations when there is none. This possibility could lead to social controversies especially when the proposed causal relations are ethically sensitive such as suggestions of racial differences of intelligence. Here again we should not accept AI's 'conclusions' automatically without human evaluation.

II.2.5. AI-based decision-making

48. AI methods can potentially have a huge impact in a wide range of areas, from the legal professions and the judiciary to aiding the decision-making of legislative and administrative public bodies. For example, they can increase the efficiency and accuracy of lawyers in both counselling and litigation, with benefits to lawyers, their clients and society as a whole. Existing software systems for judges can be complemented and enhanced through AI tools to support them in drafting new decisions (CEPEJ, 2018).

49. A key issue in such uses is the nature and interpretation of the results of algorithms, which are not always intelligible to humans[1]. This issue can be expanded to the wider field of data-driven decision-making. Being able to analyse, process and categorize very large

[1] As K.D. Ashley states: "since a Machine Learning (ML) algorithm learns rules based on statistical regularities that may surprise humans, its rules may not necessarily seem reasonable to humans. […] Although the machine-induced rules may lead to accurate predictions, they do not refer to human expertise and may not be as intelligible to humans as an expert's manually constructed rules. Since the rules the […] algorithm infers do not necessarily reflect explicit legal knowledge or expertise, they may not correspond to a human expert's criteria of reasonableness." (Ashley, 2017, p.111)

amounts of potentially rapidly-evolving data of very different natures, an AI engine is seen to be capable of proposing – and if allowed, making – decisions in complex situations. Examples of such uses discussed in this report include environmental monitoring, disaster prediction and response, anticipation of social unrest and military battlefield planning.

50. The validity of an AI-driven decision however should be treated with caution. Such a decision is not necessarily fair, just, accurate or appropriate. It is susceptible to inaccuracies, discriminatory outcomes, embedded or inserted bias and limitations of the learning process. Not only does a human have a much larger 'world view', but he or she also has a tacit knowledge that will outperform AI in critical and complex situations, such as battlefield decisions. Ideally, a decision would be the one a human would make if he or she had been able to process the mountain of data in a reasonable time. However, humans have different capabilities and make decisions based on fundamentally different decision-making architectures, including sensitivity to potential bias.

51. It is highly questionable that AI will – at least in the near future – have the capacity to cope with ambiguous and rapidly evolving data, or to interpret and execute what human intentions would have been if the human could have coped with complex and multifaceted data. Even having a human 'in the loop' to moderate a machine decision may not be sufficient to produce a 'good' decision: as cognitive AI does not make decisions in the same way as humans would, the human would not be equipped with the knowledge and information she or he would need in order to decide if the data-driven action fulfils the human's intentions. Moreover, the stochastic behaviour of cognitive AI, together with the human's consequent inability to know why a particular choice has been made by the system, means the choice is less likely to be trusted.

52. A cautionary tale that illustrates some of the problems of using AI to assist decision-making in social contexts is the Allegheny Family Screening Tool (AFST), a predictive model used to forecast child neglect and abuse in Allegheny, Pennsylvania (see https://www.alleghenycountyanalytics.us/wp-content/uploads/2017/07/AFST-Frequently-Asked-Questions.pdf). The tool was put in place with the belief that data-driven decisions would provide the promise of objective, unbiased decisions that would solve the problems of public administration with scarce resources. The Authority that implemented this tool may have been well intentioned. However, recent research has argued that the AFST tool has harmful implications for the population it hoped to serve (Eubanks, 2018b, p.190; Eubanks, 2018a). It oversamples the poor and uses proxies to understand and predict child abuse in a way that inherently disadvantages poor working families. It thus exacerbates existing structural discrimination against the poor and has a disproportionately adverse impact on vulnerable communities.

53. In some contexts, employing AI as a (either human-assisted or fully autonomous) decision maker might even be seen as a pact with the devil: in order to take advantage of the speed and large data ingestion and categorization capabilities of an AI engine, we will have to give up the ability to influence that decision. Moreover, the effects of such decisions can be profound, especially in conflict situations.

II.3. Culture and Cultural Diversity

54. AI is likely to have substantial implications for culture and artistic expression. Although still in its infancy, we are beginning to see the first instances of artistic collaboration between intelligent algorithms and human creativity, which might eventually bring important challenges for the rights of artists, the Cultural and Creative Industries (CCI), and the future of heritage. At the same time, the role of algorithms in online

streaming media and in machine translation is likely to have implications for cultural diversity and language.

II.3.1. Creativity

55. Artificial Intelligence is increasingly connected to human creativity and artistic practice: ranging from 'autotune' software that automatically corrects the pitch of the voices of singers, to algorithms helping to create visual art, compose music or write novels and poetry. Creativity, understood as the capacity to produce new and original content through imagination or invention, plays a central role in open, inclusive and pluralistic societies. For this reason, the impact of AI on human creativity deserves careful attention. While AI is a powerful tool for creation, it raises important questions about the future of art, the rights and remuneration of artists and the integrity of the creative value chain.

56. The case of the 'Next Rembrandt' – in which a brand-new Rembrandt painting was produced using AI and a 3D printer – is a good illustration (Microsoft Europe, 2016). Works of art like this require a new definition of what it means to be an 'author', in order to do justice to the creative work of both the 'original' author and of the algorithms and technologies that produced the work of art itself. This raises another question: What happens when AI has the capacity to create works of art itself? If a human author is replaced by machines and algorithms, to what extent copyrights can be attributed at all? Can and should an algorithm be recognized as an author, and enjoy the same rights as an artist?

57. Although AI is clearly capable of producing 'original' creative works, people are always involved in the development of AI technologies and algorithms, and often in the creation of artworks that serve as the inspiration for AI-generated art. From this perspective, AI can be seen as a new artistic technique, resulting in a new type of art. If we want to preserve the idea of authorship in AI creations, an analysis of the various authors 'behind' each work of art, and their relationships with each other, need to be made. Accordingly, we need to develop new frameworks to differentiate piracy and plagiarism from originality and creativity, and to recognize the value of human creative work in our interactions with AI. These frameworks are needed to avoid the deliberate exploitation of the work and creativity of human beings, and to ensure adequate remuneration and recognition for artists, the integrity of the cultural value chain, and the cultural sector's ability to provide decent jobs.

II.3.2. Cultural diversity

58. AI also has a close relation to cultural diversity. While it has the potential to positively impact the cultural and creative industries, not all artists and entrepreneurs have the skills and resources to use AI-based technologies in the creation and distribution of their work. The commercial logic of large platforms may lead to an increased concentration of cultural supply, data and income in the hands of only a few actors, with potential negative implications for the diversity of cultural expressions more generally, including the risk of creating a new creative divide, and an increasing marginalization of developing countries.

59. As these platforms develop into the dominant means of enjoying works of art, it is crucial to ensure diversity and fair access to these platforms for artists from all genres and backgrounds. In this context, artists from developing countries require special consideration. Artists and cultural entrepreneurs should have access to the training, financing opportunities, infrastructure and equipment necessary to participate in this new cultural sphere and market.

60. Moreover, the algorithms used by media streaming companies such as Spotify and Netflix have a major influence on the selection of music and movies that people enjoy. Because these platforms not only make works of art available, but also *suggest* works of art for their users to enjoy, it is important that their algorithms are designed in such a way that they do not privilege specific works of art over others by limiting their suggestions to the most dominant works of a particular genre, or to the most popular choices of users and their peers. Other institutions have expressed similar concerns (ARCEP, 2018*).* Transparency and accountability of these algorithms are essential for ensuring access to diverse cultural expressions and active participation in cultural life.

61. Also in its relation to cultural heritage, AI can play an important role. AI can be used, for instance, to monitor and analyse changes to heritage sites, in relation to development pressures, climate change, natural disasters and armed conflicts. It can also be used to monitor the illicit trafficking of cultural objects and the destruction of cultural property, and to support data collection for recovery and reconstruction efforts.

II.3.3. Language

62. In our rapidly globalizing world, the machine-powered translation of languages is likely to play an increasingly important role. Because of this, AI will have a substantial impact on language and human expression, in all dimensions of life. This fact brings with it a responsibility to deal carefully with ‘natural’ languages (as opposed to artificial languages or computer code) and their diversity. Language, after all, is the basis for human identity, social cohesion, education, and human development. Since its founding, UNESCO has recognized the importance of language in promoting access to quality education, building inclusive knowledge societies and transmitting cultural heritage and expressions (UNESCO, 2002).

63. A central element of the complex relationship between AI and language is the intermediary role of ‘formal languages’ (languages with words derived from an alphabet). AI technologies often require that words and sentences expressed in any of the many natural languages used around the world have to be translated into formal languages that can be processed by computers. The translation of many natural languages into formal languages is not a neutral process, because every translation from natural language into formal language results in the ‘loss’ of meaning, given the fact that not all the specificities and idiosyncrasies of languages can be entirely formalized.

64. A second element is the translation between natural languages, which takes place via these formal languages. There are several intrinsic problems with machine translations: words can have different meanings in different languages, and there can be a lack of linguistic or conceptual correspondence between languages. In these cases, translation is very difficult, if not technically impossible. In addition, the contextual and cultural connotations of words and expressions are not always fully translatable. Although greatly improved in recent years, at least for more common languages, automatic translation or machine translation is often too unreliable to be used, for instance, in technical fields where lexical and conceptual precision is crucial, or in cultural expression and literature.

65. These two aspects of machine translation have important implications, not only for the quality of translation and the risk of inter-language misunderstanding, but also for linguistic diversity. It is very likely that machine translation, at least in the short term, will be primarily developed for the main world languages, especially English. The technology requires large data sets compiled from human-made translations. Such data sets are often not available in significant numbers for less spoken languages. At the same time, this

technology can also play a positive role, allowing people to express themselves in less widely spoken languages.

66. An analogous process has actually already taken place with radio. While commercial radio largely produces content in widely spoken languages, thus reinforcing the cultures embodied in dominant languages, community broadcasters often generate content in local languages, thus enhancing pluralism and diversity in the media. As the UNESCO handbook on Community Media states: “[Community media is] present in all regions of the world as social movements and community-based organizations have sought a means to express their issues, concerns, cultures and languages” (UNESCO, 2013, p.7). Mass media, therefore, can actually help to preserve languages and cultural diversity.

67. Similarly, machine translation has already been used as a tool to foster diversity and protect indigenous languages. For instance, in Australia, a researcher from the ARC *Centre of Excellence for the Dynamics of Language* has recorded nearly 50,000 hours of spoken word. To process these recordings, linguists needed to select short segments of the recording that might include key sections of grammar and vocabulary, by listening to the recordings and transcribing them. Without AI, this would have taken roughly 2 million hours. So far, this usage of AI has facilitated the modelling of 12 indigenous languages spoken in Australia, including *Kunwok, Kriol, Mangarayi, Nakkara, Pitjantjatjara, Warlpiri, Wubuy*, among others (O’Brien, 2018).

68. These examples show that AI, like any technology, should be developed and used in ways that do not threaten cultural diversity but rather preserve it. If we want to preserve multilingualism and interoperability among different languages, adequate technical and financial resources should be made available to make this possible (Palfrey and Gasser, 2012; Santosuosso and Malerba, 2015).

II.4. Communication and information

69. Artificial Intelligence plays an increasingly important role in the processing, structuring and provision of information. Automated journalism and the algorithmic provision of news on social media are just a few examples of this development, raising issues of access to information, disinformation, discrimination, freedom of expression, privacy, and media and information literacy. At the same time, attention is needed for new digital divides between countries and within different social groups.

II.4.1. Disinformation

70. AI can strengthen the free flow of information and journalistic activity, but it can also be used to spread disinformation, which is sometimes referred to using the contested term ‘fake news’. Recent examples, such as the Cambridge Analytica affair, have shown that algorithms that were designed to avoid human political bias in deciding which content will appear prominently on social media can be taken advantage of for deliberately promoting the spreading of fabricated, manipulative and divisive content to specific target groups. In some cases, this content may include information fraudulently formatted as news, and may also include content that serves as emotive propaganda.

71. This can have negative effects on norms of civil and informed discussion, on social trust and public debate or even on democratic processes. The existence of different, sometimes polarized opinions is a regular feature of any open and democratic society that offers a free and open public space. Social media algorithms, however, may exacerbate the polarization of opinions by intensifying and amplifying emotional content via ‘likes’,

'shares', 'retweets', auto-completion in search queries, and other forms of online recommendations and engagement, resulting in so-called 'filter bubbles' and 'echo chambers' instead of providing an infrastructure for discussion and debate. Persons sharing the same 'bubble' may be exposed to filtered content of information and in return, the open public space can become characterized with more and more homogenized opinion groups which are at the same time more and more polarized to each other.

72. Although some big social media companies are beginning to recognize the problem and the need to address it in a multi-stakeholder way, which includes civil society together with state regulators, the solutions still seem to be unclear. One way to explore solutions is to use the UNESCO R.O.A.M. framework (Rights, Openness, Accessibility to all, Multi-stakeholder participation) to systematically identify where improvements can be made and how these interrelate with the totality of principles at stake.

73. Sometimes, the moderation of content can be justified precisely as a means to avoid spreading disinformation and content that incites violence, hatred and discrimination, as well as a means to prevent aggressive personal communication. The filtering may be done by humans, but is often assisted or even automated via AI algorithms. The particular challenge in this case is not just to identify the offending content, but also to avoid the filter being too inclusive and consequently incurring accusations of automated censorship and restriction on legitimate speech. Response to disinformation and 'hate speech' should be based on international freedom of expression standards and in line with UN conventions and declarations on the issue (Article 19, 2018a).

II.4.2. Data Journalism and Automated Journalism

74. The recent emergence of functionally powerful AI has implications for journalism in several different ways. On the one hand, the growing possibilities to use data and computer tools in journalistic research can strengthen journalistic work. On the other hand, AI might also take over some journalistic tasks. Routine tasks for which lots of 'practice data' are available are the first candidates to be mimicked by AI, and a substantial part of journalistic work is in fact routine: collecting and selecting relevant data, summarizing the results and describing them in a clear way. AI is already performing relatively simple, fixed-format jobs of article writing, in areas where continuous updates are needed, like market reviews or sports reporting. This development is ambivalent: it can also free journalists up to do higher end work in interpretation, analysis, verification and presentation of news.

75. Automated news writing without human intervention or supervision is a reality that is often hidden to the reader. As early as in 2006, some news services (e.g. Thomson Financial) announced the use of computers to generate stories based on data, in order to deliver information to their users in a fast manner. In 2014, Wibbitz (Israel) won the Netexplo Grand Prix at the UNESCO/Netexplo Forum, proposing an app that enables news channels to easily create videos using text content from the internet, providing a summary of the main ideas of the text. In recent times, a number of major mainstream media are using 'robot journalism': Le Monde, Press Association, Xinhua, to name a few, have reported to use natural language generation algorithms to cover different journalistic topics.

76. Media content production and dissemination increasingly delegate analytical and decision-making authority attributed to sophisticated algorithms. Media organizations increasingly rely on algorithms that analyse user preferences and media consumption patterns (personalization). Applied to journalism, algorithms are then called to analyse specific geographic communities for demographic, social, and political variables in order to produce the most relevant information for these communities, including weather

forecasts and sports reports. This practice has the potential to sustain local journalism and newspapers. In this way, AI can help strengthen business models for journalism.

77. At the same time, AI-based journalism raises issues of liability, transparency and copyright. Liability can be an issue when it is complicated to determine the fault in algorithm-based reporting, for instance in cases of defamation. Transparency and credibility are issues when consumers do not or cannot realize when content is machine-generated, from which sources it comes and how verified or even false the information is – with current discussions about 'deep fakes' as extreme cases. Copyright is an upcoming issue, since AI-generated content depends ever less on human input, which is reason for some to argue that some form of copyright liability should be attributed to the algorithms themselves.

78. To address these challenges, many argue that journalists and editors should engage with the technologists who build the algorithms. An example of this is the recent launch of an open-source platform by Quartz AI Studio, a US-based project to help journalists use machine learning supporting them in various tasks.

II.5. AI in Peace-Building and Security

79. In line with UNESCO's mission and mandate to promote and build peace, this study also wants to investigate the role of Artificial Intelligence in matters of peace-building and security. The fact that this includes the potential military use of AI in no way weakens its commitment to peace.

80. AI is argued to be capable of analysing, processing and categorizing very large amounts of rapidly evolving data of very different natures (Payne, 2018; Roff, 2018; Gupta, 2018). '*Hard'* data would include satellite and other surveillance imagery, signals and electronic intelligence, while *'soft'* data could include reports, documents, newsfeeds, social media inputs and political and sociological data. AI is advertised as being capable of categorizing this massive amount of data to identify external and internal threats, discover the objectives and strategies of actors, interpret complex and multifaceted intentions underpinning their activities, and strategies about how to pre-empt or counter predicted actions.

81. Such a situational awareness tool could be a powerful instrument for conflict prevention and resolution (Spiegeleire et al., 2017). It could give insight into the drivers of human endeavour and their outcomes, with possible application in deradicalization. Learning-enabled 'anticipatory intelligence' might foresee the development of social unrest and societal instability, and suggest means of prevention. Deeper insights into the drivers of conflict could nudge potential agents of conflict away from realizing malign intentions. We might be able to detect social pathologies at an early stage, find out which actions might de-escalate a threatening situation, or discover effective non-inflammatory routes to counter attempts to whip up sectarian frenzy. At a societal level, by tracking and helping us understand the dynamics that strengthen or weaken societal resilience, AI may be able to lead us to a more resilient society, and to help us move towards a more peaceful, conflict-free world.

82. On the negative side, *AI will transform the nature and practice of conflict,* with a consequential impact on society that will reach far beyond strictly military issues (Payne, 2018; Spiegeleire et al., 2017). Not only will it change how explosive force is used by increasing the effectiveness of deployment of weapons systems, but also AI promises to dramatically improve the speed and accuracy of everything from military logistics, intelligence and situational awareness to battlefield planning and execution/operations.

The AI system itself might be used to make its own suggestions of actions to be taken: it could create a set of orders that exploit enemy weaknesses that it has identified from its own analysis, or, from finding patterns in enemy/insurgent actions, devise countermeasures to predicted aggressive action. It might also do its own 'war gaming' to probe the likely responses to particular actions.

83. The speed with which such planning tools could operate would increase the ability to act under rapidly changing situations. One can envisage, for example, the development of algorithmic response to coordinated attack by e.g. drone swarms and other uninhabited assets such as incoming missiles. The speed of AI-enabled response can be seen as an incentive to use it and hence be potentially destabilising. Or indeed disastrous, as past examples of machine warnings being thankfully not being acted on by an intervening human commander have demonstrated. Nevertheless, a State that does not go down this AI response route would be at a major disadvantage, thus encouraging proliferation of the capability.

84. The possibility exists of the AI-assisted decision-making machine implementing its own attack and kill decisions without human intervention – for example, a fully autonomous weapon. The idea of such a *non-human* entity having specific agency could radically change our understanding of politics at the widest levels. Moreover, the closeness of potential military uses of AI to its civilian development ('ease of weaponisation') means it is not a discretely bounded category, a characteristic which complicates both the ethics and the regulation of its development and application.

85. While AI might be considered to be just another revolution in military affairs that allows armed forces to do similar things with similar tools, perhaps its real 'revolutionary' potential (Payne, 2018; Spiegeleire et al., 2017) is in *transforming the concept of 'armed force'* into one whose weapons are more subtle than explosive devices. The power of AI in conflict lies not only in enhancing physical technologies, but also in redefining what 'armed force' might be.

86. We are already seeing this in the cyber context, where AI gives it both defence and attack capability. Through pattern matching, deep learning and observing deviations from normal activity, software vulnerabilities can be detected and then weaponized to avoid defences. Deep neural networks may detect and prevent intrusions. In order to be effective, cyber defences will have to operate at speed and by implication have a high degree of autonomy.

87. Propaganda is another weapon that AI has empowered. The ease of faking voices, images and news, and propagating them to selected audiences, threatens social engineering and (mis)-shaping of public opinion. In essence, AI makes it easier to lie persuasively and enhance forgery. The consequent threat to trust in the integrity of information increases the potential for miscalculation of a perceived adversary's intention both tactically and strategically.

88. AI also empowers economic sabotage and critical infrastructure disruption. By moving radio and electronic warfare into cognitive mode, AI could be critical in interfering access to the electromagnetic spectrum. Systems are already marketed, which use machine learning, 'intelligent' algorithms and adaptive signal processing.

89. Finally, with respect to *internal* state security, the use of data set analysis and face recognition implies a new relationship between society and the institutions charged with protecting it. This obviously has significant ethical implications.

II.6. AI and Gender Equality

90. AI systems have significant implications for gender equality, since they may reflect existing societal biases, with the potential to exacerbate them. Most AI systems are built using datasets that reflect the real world – one which can be flawed, unfair, and discriminatory (Marda, 2018). Recently, a hiring tool used by Amazon was found to be sexist, as it prioritized male applicants for technical jobs (Reuters, 2018). Such systems can be dangerous, not only because they perpetuate gender inequalities in society, but also because they embed these inequalities in opaque ways, while at the same time being hailed as 'objective' and 'accurate' (O'Neil, 2018).

91. These inequalities are primarily a result of how machines learn. Indeed, as machine learning relies on the data it is fed, particular attention is needed to promote gender-sensitive data as well as gender-disaggregated data collection. In the case of Amazon's hiring tool, the bias emerged because the tool was learning from previous Amazon candidates – who were predominantly male – and had 'learnt' that male applicants should be preferred over female applicants (Short, 2018). Paying attention to biased data would therefore help limit the blindspot of how AI systems can best be suited and designed for both men and women. Additionally, applying gender-disaggregated data to AI analytics represents an opportunity to better grasp gender issues we face today.

92. It is important to note that gender inequalities begin at the early stages of conceptualising and designing AI systems. The gender disparity in technical fields is well known and apparent (Hicks, 2018), from wage gaps to promotions (Brinded, 2017). This is generally known as the 'leaky pipeline', with female participation in tech and engineering dropping 40% between when students graduate, to when they become executives in the field (Wheeler, 2018). The low share of women in the AI workforce – and in digital skills development in general – means that women's voices are not equally represented in the decision-making processes that go into the design and development of AI systems. As a result, we risk building these technologies only for some demographics (Crawford, 2016).

93. Moreover, the biases that people carry in their everyday lives can be reflected and even amplified through the development and use of AI systems. The 'gendering' of digital assistants, for example, may reinforce understandings of women as subservient and compliant. Indeed, female voices are routinely chosen as personal assistance bots, mainly fulfilling customer service duties, whilst the majority of bots in professional services such as the law and finance sectors, for example, are coded as male voices. This has educational implications with regards to how we understand 'male' vs 'female competences, and how we define authoritative versus subservient positions. Further, the notion of 'gender' in AI systems is often a simple choice - male or female. This ignores and actively excludes transgender individuals, and can discriminate against them in humiliating ways (Costanza-Chock, 2018).

II.7. Africa and AI Challenges

94. Africa, like other developing regions, is facing the acceleration of the use of information technologies and AI. The new digital economy that is emerging presents important societal challenges and opportunities for African creative societies.

95. Concretely, in terms of infrastructure connectivity, Africa has a very large deficit and is significantly behind other developing regions; domestic connections, regional links and continuous access to electricity are a big handicap. Infrastructure services are paid at

a high price even if more and more Africans - even in town's slums - have their own mobile phones.

96. The development issues that African countries face are numerous. The human rights framework and the Sustainable Development Goals (SDGs) provide a consistent way to orient the development of AI. Therefore, how can AI technology and knowledge be shared and oriented through the priorities defined by developing countries themselves? These include challenges such as infrastructure, skills, knowledge gaps, research capacities and availability of local data, as expressed during the UNESCO Forum on Artificial Intelligence in Africa that took place at the Mohammed VI Polytechnic University, in Benguerir, Morocco, on 12 and 13 December 2018.

97. The role of women is crucial. As very dynamic economic agents in Africa, women carry out the majority of agricultural activities, hold one-third of all businesses and may represent, in some countries, up to 70% of employees. They are the main levers of the domestic economy and family welfare, and play an absolutely indispensable leadership role in their respective communities and nations. By placing gender equality at the centre of its strategy for promoting development in Africa, the African Development Bank recognizes the fundamental role of gender parity in achieving inclusive growth and in the emergence of resilient societies. Access to education, to AI literacy and more globally to information and communication technologies (ICTs) are key elements to empower women in order to avoid their marginalization.

98. With particular attention to scientific research, science, technology, engineering and mathematics, together with education for citizenship based on values, rights and obligations, AI should be integrated into national development policies and strategies by drawing on endogenous cultures, values and knowledge in order to develop African economies.

III. STANDARD-SETTING INSTRUMENT

III.1. Declaration vs Recommendation

99. The Working Group carefully examined two of UNESCO's normative tools – the Declaration and the Recommendation –, which were linked to the analyses of the two first sections of this preliminary study on the Ethics of AI. The Working Group also drew on COMEST's previous experience, which initiated the 2017 Declaration of Ethical Principles in relation to Climate Change and participated in the revision of the 2017 Recommendation on Science and Scientific Researchers. The Working Group weighed the pros and cons of each of these two normative tools.

100. With regard to the proposal for a Declaration on the Ethics of Artificial Intelligence, the Working Group noted the very recent increase in the number of declarations of ethical principles on AI in 2018. The *Montreal Declaration for a Responsible Development of AI* (University of Montreal, 2018), the *Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems* (Amnesty International and Access Now, 2018), and the Declaration of the Future of Life Institute on the *Asilomar AI Principles* (Future of Life Institute, 2017) come from different initiatives and are supported by various organizations (universities, governments, professional associations, companies, NGOs). We must add to this set of declarations several ethical proposals such as: the *Ethics Guidelines for Trustworthy AI* from the European Commission's High-Level Expert Group on AI, which is based on Human Rights; and the second document of the IEEE (currently under consultation) on *Ethically Aligned Design: A Vision for Prioritizing Human Well-being*

with Autonomous and Intelligent Systems, which is addressed to engineers and aimed to embed values into Autonomous Intelligent Systems. All these initiatives are positive as they initiate discussions on AI ethics at different levels.

101. Nevertheless, the Working Group concluded that there was a great heteronomy in the principles and in the implementation of the values promoted by one or the other. This heteronomy is both the consequence of the definition that has been chosen for AI, and of the objectives that are being sought - governance, education for engineers, public policy. The question is as follows: would a UNESCO Declaration on the Ethics of AI allow this heteronomy to be federated under a few guiding principles that would respond in a comprehensive manner to the ethical issues of AI, as well as to UNESCO's specific concerns in the fields of education, culture, science and communication? The Working Group believes that this could be possible, but with the risk that during the process leading to the Declaration, Member States would essentially agree on some general, abstract and non-binding principles, since it is a Declaration. In such a perspective, would a UNESCO Declaration on the Ethics of AI bring added value vis-a-vis other ongoing declarations and initiatives? It is questionable that such instrument will immediately establish itself as an international reference, in a context of competition between ethical frameworks, at a time when technologies are emerging and their uses not yet stabilized.

102. The Working Group therefore considered whether a Recommendation would then be a more appropriate tool in the current situation. At the international level, the European level and the national political context for several countries, there is a move towards similar forms of regulation with respect to the digital economy, but also taking into account the relations between the two major digital powers - USA and China. The increase of criticisms concerning the non-transparency, biases or ways of acting by big companies, or the rise of popular mistrust in the face of cyber-attacks are creating a new political climate that is having an impact on the development of AI. The digital regulation movement, initiated by the European Union on the protection of personal data, could therefore be extended to an international level in emerging fields such as AI. However, at this level, the tools are still in their early stages of development, although the OECD's strategy through its Artificial Intelligence Expert Group (AIGO) emphasizes responsibility, security, transparency, protection and accountability:

> The OECD supports governments through policy analysis, dialogue and engagement and identification of best practices. We are putting significant efforts into work on mapping the economic and social impacts of AI technologies and applications and their policy implications. This includes improving the measurement of AI and its impacts, as well as shedding light on important policy issues such as labour market developments and skills for the digital age, privacy, accountability of AI-powered decisions, and the responsibility, security and safety questions that AI generates. (OECD, 2019)

103. OECD public policy priorities are more a matter of AI governance and good practice. It seems here that UNESCO's approach could be complementary at the international level to the OECD’s, but with a focus on aspects that are generally neglected such as culture, education, science and communication. These dimensions directly affect people and populations in their daily lives and in their individual and collective aspirations. UNESCO's approach for a Recommendation on AI Ethics would be presented as a complementary alternative to a vision of economic governance. The Working Group therefore believes that by initiating a Recommendation, although it requires more time and energy than a Declaration, UNESCO would be able to distinguish itself not only in terms of ethical content but also through specific proposals to Member States. One of the aims

is to empower and strengthen the capacity of States to intervene in key areas that are impacted by the development of AI, such as culture, education, science and communication.

104. The Recommendation should contain two dimensions. The first is the affirmation of a number of basic principles for an Ethics of AI. The second is the outlining of specific proposals to help States monitor, and regulate the uses of AI in the areas under UNESCO's mandate through the reporting mechanism of the Recommendation, as well as identify ethical assessment tools to review on a regular basis their policies for guiding the development of AI. In this regard, UNESCO would be uniquely positioned to provide a multidisciplinary perspective, as well as a universal platform for the development of a Recommendation on the Ethics of AI. Specifically, UNESCO would be able to bring together both developed and developing countries, different cultural and moral perspectives, as well as various stakeholders within the public and private spheres into a truly international process for elaborating a comprehensive set of principles and proposals for the Ethics of AI.

105. The next section identifies some of these proposals.

III.2. Suggestions for a standard-setting instrument

106. On the basis of its analysis of the potential implications of Artificial Intelligence for society, the Working Group would like to suggest a number of elements that could be included in an eventual Recommendation on the Ethics of AI. These suggestions embody the global perspective of UNESCO, as well as UNESCO's specific areas of competence.

107. First of all, the Working Group would like to suggest a number of generic principles for the development, implementation and use of AI. These principles are:

a. **Human rights**: AI should be developed and implemented in accordance with international human rights standards.
b. **Inclusiveness**: AI should be inclusive, aiming to avoid bias and allowing for diversity and avoiding a new digital divide.
c. **Flourishing**: AI should be developed to enhance the quality of life.
d. **Autonomy**: AI should respect human autonomy by requiring human control at all times.
e. **Explainability**: AI should be explainable, able to provide insight into its functioning.
f. **Transparency**: The data used to train AI systems should be transparent.
g. **Awareness and literacy**: Algorithm awareness and a basic understanding of the workings of AI are needed to empower citizens.
h. **Responsibility**: Developers and companies should take into consideration ethics when developing autonomous intelligent system.
i. **Accountability**: Arrangements should be developed that will make possible to attribute accountability for AI-driven decisions and the behaviour of AI systems.
j. **Democracy**: AI should be developed, implemented and used in line with democratic principles.
k. **Good governance**: Governments should provide regular reports about their use of AI in policing, intelligence, and security.

l. **Sustainability**: For all AI applications, the potential benefits need to be balanced against the environmental impact of the entire AI and IT production cycle.

108. More specifically, the Working Group would like to point out some central ethical concerns regarding the specific focus of UNESCO:

a. **Education**: AI requires that education fosters AI literacy, critical thinking, resilience on the labour market, and educating ethics to engineers.

b. **Science**: AI requires a responsible introduction in scientific practice, and in decision-making based on AI systems, requiring human evaluation and control, and avoiding the exacerbation of structural inequalities.

c. **Culture**: AI should foster cultural diversity, inclusiveness and the flourishing of human experience, avoiding a deepening of the digital divide. A multilingual approach should be promoted.

d. **Communication and information**: AI should strengthen freedom of expression, universal access to information, the quality of journalism, and free, independent and pluralistic media, while avoiding the spreading of disinformation. A multi-stakeholder governance should be promoted.

e. **Peace**: In order to contribute to peace, AI could be used to obtain insights in the drivers of conflict, and should never operate out of human control.

f. **Africa**: AI should be integrated into national development policies and strategies by drawing on endogenous cultures, values and knowledge in order to develop African economies.

g. **Gender**: Gender bias should be avoided in the development of algorithms, in the datasets used for their training, and in their use in decision-making.

h. **Environment**: AI should be developed in a sustainable manner taking into account the entire AI and IT production cycle. AI can be used for environmental monitoring and risk management, and to prevent and mitigate environmental crises.

BIBLIOGRAPHY

AI Now. 2016. *The AI Now Report: The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term*. New York, The White House and the New York University's Information Law Institute. Available at: https://ainowinstitute.org/AI_Now_2016_Report.pdf

Ajunwa, I., Crawford, K., and Schultz, J. 2017. Limitless Worker Surveillance. *California Law Review.* No. 735, pp. 101-142.

Allen, G. and Chan, T. 2017. Artificial Intelligence and National Security. *Harvard Kennedy School, Belfer Center for Science and International Affairs.* Online. Available at: https://www.belfercenter.org/publication/artificial-intelligence-and-national-security

Amnesty International and Access Now. 2018. *The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems.* Toronto, RightsCon 2018. Available at: https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf

ARCEP (Autorité de régulation des communications électroniques et des postes). 2018. *Smartphones, tablets, voice assistants... Devices, the weak link in achieving an open Internet*. Paris, ARCEP. Available at: https://www.arcep.fr/uploads/tx_gspublication/rapport-terminaux-fev2018-ENG.pdf

Article 19. 2018a. *Free speech concerns amid the "fake news" fad*. Online. Available at: https://www.article19.org/resources/free-speech-concerns-amid-fake-news-fad/

Article 19. 2018b. *Privacy and Freedom of Expression in the Age of Artificial Intelligence*. Online. Online. Available at: https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf

Ashley, K.D. 2017. Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age. Cambridge, Cambridge University Press.

Boden, M.A. 2016. *AI: Its Nature and Future*. Oxford, Oxford University Press.

Brinded, L. 2017. "Robots are going to turbo charge one of society's biggest problems", *QUARTZ* (28 December 2017). Online. Available: https://qz.com/1167017/robots-automation-and-ai-in-the-workplace-will-widen-pay-gap-for-women-and-minorities/

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B. and Anderson, H. 2018. *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. Available at: https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf

Bunnin, N. and Yu, J. 2008. *The Blackwell dictionary of western philosophy*. John Wiley & Sons.

Butterfield, A., Ngondi, G.E. and Kerr, A. eds. 2016. *A dictionary of Computer Science*. Oxford, Oxford University Press.

Costanza-Chock, S. 2018. “Design justice, AI, and escape from the matrix of domination”, *Journal of Design and Science.* Online. Available at: https://jods.mitpress.mit.edu/pub/costanza-chock

Crawford, K. 2016. “Artificial Intelligence’s White Guy Problem”, *The New York Times* (Opinion, 25 June 2016). Online. Available at: https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html

Crawford, K. 2017. ‘The Trouble with Bias’, NIPS 2017 Keynote. Available at: https://www.youtube.com/watch?v=fMym_BKWQzk

Cummings, M. L., Roff, H. M., Cukier, K., Patakilas, J. and Bryce, H. 2018. *Artificial Intelligence and International Affairs: Disruption Anticipated.* Chatham House Report. Available at: https://www.chathamhouse.org/sites/default/files/publications/research/2018-06-14-artificial-intelligence-international-affairs-cummings-roff-cukier-parakilas-bryce.pdf

Brookfield Institute and Policy Innovation Hub (Ontario). 2018. *Policymakers: Understanding the Shift*. Online. Available at: https://brookfieldinstitute.ca/wp-content/uploads/Brookfield-Institute_-The-AI-Shift.pdf

Eubanks, V. 2018a. “A Child Abuse Prediction Model Fails Poor Families”, *WIRED*. Online. Available at: https://www.wired.com/story/excerpt-from-automating-inequality/

Eubanks, V. 2018b. *Automating Inequality: How high tech tools profile, police, and punish the poor*. New York, St. Martin’s Press.

European Commission (EC). 2018. *Artificial Intelligence for Europe.* Communication from the Commission to the European Parliament, the European council, the Council, the European Economic and Social Committee and the Committee of the Regions. Brussels, European Commission. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237&from=EN

European Commission for the Efficiency of Justice (CEPEJ). 2018. *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment.* Strasbourg, CEPEJ. Available at: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c

European Group on Ethics in Science and New Technologies (EGE). 2018. *Statement on AI, Robotics, and Autonomous System*. Brussels, European Commission. Available at: https://publications.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1/language-en/format-PDF/source-78120382

Executive Office of the President (USA). 2016. *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*. Washington, D.C., Executive Office of the President. Available at: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf

Frankish, K. and Ramsey, W.M. eds. 2014. *The Cambridge handbook of artificial intelligence*. Cambridge, Cambridge University Press.

Future of Life Institute. 2017. *Asilomar AI Principles.* Cambridge, Future of Life Institute. Available at: https://futureoflife.org/ai-principles/?cn-reloaded=1

Gupta, D.K. 2018. "Military Applications of Artificial Intelligence", *Indian Defence Review* (22 March 2019). Online. Available at: http://www.indiandefencereview.com/military-applications-of-artificial-intelligence/

Heacock, M., Kelly, C.B., Asante, K.A., Birnbaum, L.S., Bergman, Å.L., Bruné, M.N., Buka, I., Carpenter, D.O., Chen, A., Huo, X. and Kamel, M. 2015. "E-waste and harm to vulnerable populations: a growing global problem", *Environmental health perspectives*, Vol. 124, No. 5, pp. 550-555.

Hicks, M. 2018. "Why tech's gender problem is nothing new", *The Guardian* (12 October 2018). Online. Available at: https://amp.theguardian.com/technology/2018/oct/11/tech-gender-problem-amazon-facebook-bias-women?__twitter_impression=true

Hinchliffe, T. 2018. "Medicine or poison? On the ethics of AI implants in humans", *The Sociable*. Online. Available at: https://sociable.co/technology/ethics-ai-implants-humans/

House of Lords. 2017. *AI in the UK: ready, willing and able?* London, House of Lords Select Committee on Artificial Intelligence. Available at: https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

Illanes, P., Lund, S., Mourshed, M., Rutherford, S. and Tyreman, M. 2018. *Retraining and reskilling workers in the age of automation*. Online, McKinsey Global Institute. Available at: https://www.mckinsey.com/featured-insights/future-of-work/retraining-and-reskilling-workers-in-the-age-of-automation

Institute of Electrical and Electronic Engineers (IEEE). 2018. *Ethically Aligned Design – Version 2 for Public Discussion*. New Jersey, The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Available at: https://ethicsinaction.ieee.org/

Laplante, P.A. 2005. *Comprehensive dictionary of electrical engineering*. Boca Raton, CRC Press.

Latonero, M. 2018. *Governing Artificial Intelligence: Upholding Human Rights & Dignity*. Data & Society. Available at: https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf

Marda, V. 2018. "Artificial Intelligence Policy in India: A Framework for Engaging the Limits of Data-Driven Decision-Making", *Philosophical Transactions A: Mathematical, Physical and Engineering Sciences*. Online. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3240384

Matias, Y. 2018. Keeping people safe with AI-enabled flood forecasting. *The Keyword* (24 September 2018). Online. Available at: https://www.blog.google/products/search/helping-keep-people-safe-ai-enabled-flood-forecasting/

Matsumoto, D.E. 2009. *The Cambridge dictionary of psychology.* Cambridge, Cambridge University Press.

McCarthy, J., Minsky, M. L., Rochester, N., Shannon, C. E. 2006 [1955]. "A proposal for the Dartmouth Summer Research Project on Artificial Intelligence", *AI Magazine*, vol. 27, no. 4, pp.12-14.

Microsoft Europe. 2016. "The Next Rembrandt", *Microsoft News Centre Europe*. Online. Available at: https://news.microsoft.com/europe/features/next-rembrandt/

National Science and Technology Council (USA). 2016. *The National Artificial Intelligence Research and Development Strategic Plan*. Washington, D.C., National Science and Technology Council. Available at: https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf

O'Brien, A. 2018. "How AI is helping preserve Indigenous languages", *SBS News*. Online. Available at: https://www.sbs.com.au/news/how-ai-is-helping-preserve-indigenous-languages

O'Neil, C. 2018. "Amazon's Gender-Biased Algorithm Is Not Alone", *Bloomberg Opinion* (16 October 2018). Online. Available at: https://www.bloomberg.com/opinion/articles/2018-10-16/amazon-s-gender-biased-algorithm-is-not-alone

OECD. 2019. *Going Digital.* Paris, OECD. Available at: http://www.oecd.org/going-digital/ai/

Oppenheimer, A. 2018. *¡Sálvese quien pueda!: El futuro del trabajo en la era de la automatización*. New York, Vintage Espanol.

Palfrey, J.G. and Gasser, U. 2012. *Interop: The Promise and Perils of Highly Interconnected Systems*. New York, Basic Books.

Payne, K. 2018. "Artificial Intelligence: A Revolution in Strategic Affairs?", *Survival*, Vol. 60, No. 5, pp. 7-32.

Peiser, J. 2019. "The Rise of the Robot Reporter", *The New York Times* (5 February 2019). Online. Available at: https://www.nytimes.com/2019/02/05/business/media/artificial-intelligence-journalism-robots.html

Reuters. 2018. "Amazon ditched AI recruiting tool that favored men for technical jobs", *The Guardian* (11 October 2018). Online. Available at: https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine

Roff, H.M. 2018. 2018. "COMPASS: a new AI-driven situational awareness tool for the Pentagon?", *Bulletin of the Atomic Scientists* (10 May 2018). Online. Available at: https://thebulletin.org/2018/05/compass-a-new-ai-driven-situational-awareness-tool-for-the-pentagon/

Rosenberg, J.M. 1986. *Dictionary of artificial intelligence and robotics*. New York, John Wiley & Sons.

Russell, S.J. and Norvig, P. 2016. *Artificial Intelligence: A Modern Approach*, 3rd ed. Harlow, Pearson.

Santosuosso, A. and Malerba, A. 2015. "Legal Interoperability As a Comprehensive Concept in Transnational Law", *Law, Innovation and Technology*, Vol. 5, No. 1, pp. 51-73.

Short, E. 2018. "It turns out Amazon's AI hiring tool discriminated against women", *Siliconrepublic* (11 October 2018). Online. Available at: https://www.siliconrepublic.com/careers/amazon-ai-hiring-tool-women-discrimination

Spiegeleire, S. De, Maas, M. and Sweijs, T. 2017. *Artificial Intelligence and the Future of Defence*. The Hague, The Hague Centre for Strategic Studies.

UNI Global Union. 2016. Top 10 principles for ethical artificial intelligence. Switzerland, UNI Global Union. Available at: http://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf

UNICEF. 2017. *Children in a Digital World*. New York UNICEF. Available at: https://www.unicef.org/publications/files/SOWC_2017_ENG_WEB.pdf

United Nations Educational, Scientific and Cultural Organization (UNESCO). 2002. *UNESCO Universal Declaration on Cultural Diversity: a vision, a conceptual platform, a pool of ideas for implementation, a new paradigm*. Paris, UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000127162

UNESCO. 2013. *Community Media: A Good Practice Handbook*. Paris, UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000215097

UNESCO. 2015a. *Keystones to foster inclusive knowledge societies: access to information and knowledge, freedom of expression, privacy and ethics on a global internet.* Paris, UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000232563

UNESCO. 2015b. *Outcome document of the "CONNECTing the Dots: Options for Future Action" Conference*. Paris, UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000234090

University of Montreal. 2018. *Montreal Declaration for a Responsible Development of AI*. Montreal, University of Montreal. Available at: https://www.montrealdeclaration-responsibleai.com/

Vernon, D. 2014. *Artificial cognitive systems: A primer*. Cambridge, MIT Press.

Villani, C., Schoenauer, M., Bonnet, Y., Berthet, C., Cornut, A.-C., Levin, F. and Rondepierre, B. 2018. *For A Meaningful Artificial Intelligence: Towards a French and European Strategy*. Paris. Available at: https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

Wheeler, T. 2018. ”Leaving at Lightspeed : the number of senior women in tech is decreasing”, *OECD Forum* (23 March 2018). Online. Available: https://www.oecd-forum.org/users/91062-tarah-wheeler/posts/31567-leaving-at-lightspeed-the-number-of-senior-women-in-tech-is-decreasing

World Summit on the Information Society (WSIS). 2003. *Declaration of Principles. Building the Information Society: A global challenge in the new Millenium*. Geneva, WSIS. Available at: http://www.itu.int/net/wsis/docs/geneva/official/dop.html

WSIS. 2005. *Tunis Agenda for the Information Society*. Tunis, WSIS. Available at: http://www.itu.int/net/wsis/docs2/tunis/off/6rev1.html

ANNEX: COMPOSITION OF THE COMEST EXTENDED WORKING GROUP ON ETHICS AND AI

1. **Prof. (Mr) Peter-Paul VERBEEK (Co-Coordinator)**
   Professor of Philosophy of Technology at the University of Twente, Netherlands
   Member of COMEST (2016-2019)

2. **Prof. (Mrs) Marie-Hélène PARIZEAU (Co-Coordinator)**
   Professor, Faculty of Philosophy, Université Laval, Québec, Canada
   Member of COMEST (2012-2019)
   Chairperson (2016-2019) and Vice-Chairperson (2014-2015) of COMEST

3. **Prof. (Mr) Tomislav BRACANOVIĆ**
   Research Associate, Institute of Philosophy, Zagreb, Croatia
   Member of COMEST (2014-2021)
   Rapporteur of COMEST (2018-2019)

4. **Mr John FINNEY**
   Emeritus Professor of Physics, Department of Physics and Astronomy, London, United Kingdom
   Coordinator of the Working Group on Scientific Ethics, Pugwash Conference on Science and World Affairs
   Ex-officio Member of COMEST

5. **Mr Javier JUAREZ MOJICA**
   Commissioner, Board of the Federal Telecommunications Institute of Mexico, Mexico City, Mexico
   Member, OECD Expert Group on AI (AIGO)
   Member of COMEST (2018-2021)

6. **Mr Mark LATONERO**
   Research Lead, Data and Human Rights, Data & Society

7. **Ms Vidushi MARDA**
   Digital Programme Officer at ARTICLE 19

8. **Prof. (Ms) Hagit MESSER-YARON**
   Professor of Electrical Engineering and former Vice-President for Research and Development, University of Tel Aviv, Tel Aviv, Israel
   Member, Executive Committee, The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems
   Member of COMEST (2016-2019)

9. **Dr (Mr) Luka OMLADIC**
   Lecturer, University of Ljubljana, Ljubljana, Slovenia
   Member of COMEST (2012-2019)

10. Prof. (Mrs) Deborah OUGHTON
Professor and Research Director, Centre for Environmental Radioactivity, Norwegian University of Life Sciences
Member of COMEST (2014-2021)

11. Prof. (Mr) Amedeo SANTOSUOSSO
Founder and Scientific Director, European Center for Law, Science and new Technologies (ECLT), University of Pavia, Pavia, Italy,
President, First Chamber, Court of Appeal of Milan, Italy
Member of COMEST (2018-2021)

12. Prof. (Mr) Abdoulaye SENE
Environmental sociologist, Coordinator for “Ethics, Governance, Environmental and Social Responsibility”, Environmental Sciences Institute, Cheikh Anta Diop University, Dakar, Senegal
Member of COMEST (2012-2019)
Vice-Chairperson of COMEST (2016-2019)

13. Prof. (Mr) John SHAWE-TAYLOR
UNESCO Chair in Artificial Intelligence, University College of London and Chair of the Knowledge 4 All Foundation

14. Mr Davide STORTI
Programme Specialist, Section for ICT in Education, Science and Culture, Communication and Information Sector, UNESCO

15. Prof. (Mr) Sang Wook YI
Professor of Philosophy, Hanyang University, Seoul, Republic of Korea
Member of COMEST (2018-2021)

United Nations
Educational, Scientific and
Cultural Organization

Ad Hoc Expert Group (AHEG) for the preparation of a draft text of a recommendation on the ethics of artificial intelligence

Distribution: limited

SHS/BIO/AHEG-AI/2020/4 REV.2
Paris, 7 September 2020
Original: English

OUTCOME DOCUMENT:

FIRST DRAFT OF THE RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE

In line with the decision of UNESCO's General Conference at its 40th session (40 C/Resolution 37), the Director-General constituted the Ad Hoc Expert Group (AHEG) for the preparation of a draft text of a recommendation on the ethics of artificial intelligence in March 2020.

Adapting to the challenging situation posed by the COVID-19 pandemic, the AHEG worked virtually from the end of March until beginning of May 2020, and produced the first version of a draft text of the Recommendation on the Ethics of Artificial Intelligence.

An extensive multi-stakeholder consultation process on this first version was conducted from June to August 2020, based on three components: (i) public online consultation (receiving more than 800 responses); (ii) regional and subregional virtual consultations co-organized with host countries/institutions in all of UNESCO's regions (involving more than 500 participants); and (iii) open, multi-stakeholder, and citizen deliberation workshops organized by partners (involving approximately 500 participants). The consultation process generated more than 50,000 comments on the text.

Taking into account the feedback received during this consultation process, the AHEG revised the first version of the draft text from August until beginning of September 2020 to produce the first draft of the Recommendation contained in this document, which will be transmitted to Member States for written comments in September 2020.

The AHEG was supported by the Assistant Director-General for Social and Human Sciences, and the Bioethics and Ethics of Science Section.

This document does not claim to be exhaustive and does not necessarily represent the views of the Member States of UNESCO.

FIRST DRAFT OF THE RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE

PREAMBLE

The General Conference of the United Nations Educational, Scientific and Cultural Organization (UNESCO), meeting in Paris from xx to xx, at its xx session,

Recognizing the profound and dynamic impact of artificial intelligence (AI) on societies, ecosystems, and human lives, including the human mind, in part because of the new ways in which it influences human thinking, interaction and decision-making, and affects education, human, social and natural sciences, culture, and communication and information,

Recalling that, by the terms of its Constitution, UNESCO seeks to contribute to peace and security by promoting collaboration among nations through education, the sciences, culture, and communication and information, in order to further universal respect for justice, for the rule of law and for the human rights and fundamental freedoms which are affirmed for the peoples of the world,

Convinced that the standard-setting instrument presented here, based on international law and on a global normative approach, focusing on human dignity and human rights, as well as gender equality, social and economic justice, physical and mental well-being, diversity, interconnectedness, inclusiveness, and environmental and ecosystem protection can guide AI technologies in a responsible direction,

Considering that AI technologies can be of great service to humanity but also raise fundamental ethical concerns, for instance regarding the biases they can embed and exacerbate, potentially resulting in inequality, exclusion and a threat to cultural, social and ecological diversity and social or economic divides; the need for transparency and understandability of the workings of algorithms and the data with which they have been trained; and their potential impact on human dignity, human rights, gender equality, privacy, freedom of expression, access to information, social, economic, political and cultural processes, scientific and engineering practices, animal welfare, and the environment and ecosystems,

Recognizing that AI technologies can deepen existing divides and inequalities in the world, within and between countries, and that justice, trust and fairness must be upheld so that no one should be left behind, either in enjoying the benefits of AI technologies or in the protection against their negative implications, while recognizing the different circumstances of different countries and the desire of some people not to take part in all technological developments,

Conscious of the fact that all countries are facing an acceleration of the use of information and communication technologies and AI technologies, as well as an increasing need for media and information literacy, and that the digital economy presents important societal, economic and environmental challenges and opportunities of benefits sharing, especially for low- and middle-income countries (LMICs), including but not limited to least developed countries (LDCs), landlocked developing countries (LLDCs) and small island developing States (SIDS), requiring the recognition, protection and promotion of endogenous cultures, values and knowledge in order to develop sustainable digital economies,

Recognizing that AI technologies have the potential to be beneficial to the environment and ecosystems but in order for those benefits to be realized, fair access to the technologies is required without ignoring but instead addressing potential harms to and impact on the environment and ecosystems,

Noting that addressing risks and ethical concerns should not hamper innovation but rather provide new opportunities and stimulate new and responsible practices of research and innovation that anchor AI technologies in human rights, values and principles, and moral and ethical reflection,

Recalling that in November 2019, the General Conference of UNESCO, at its 40th session, adopted 40 C/Resolution 37, by which it mandated the Director-General "to prepare an international standard-setting instrument on the ethics of artificial intelligence (AI) in the form of a recommendation", which is to be submitted to the General Conference at its 41st session in 2021,

Recognizing that the development of AI technologies results in an increase of information which necessitates a commensurate increase in media and information literacy as well as access to critical sources of information,

Observing that a normative framework for AI technologies and its social implications finds its basis in ethics, as well as human rights, fundamental freedoms, access to data, information and knowledge, international and national legal frameworks, the freedom of research and innovation, human and environmental and ecosystem well-being, and connects ethical values and principles to the challenges and opportunities linked to AI technologies, based on common understanding and shared aims,

Recognizing that ethical values and principles can powerfully shape the development and implementation of rights-based policy measures and legal norms, by providing guidance where the ambit of norms is unclear or where such norms are not yet in place due to the fast pace of technological development combined with the relatively slower pace of policy responses,

Convinced that globally accepted ethical standards for AI technologies and international law, in particular human rights law, principles and standards can play a key role in harmonizing AI-related legal norms across the globe,

Recognizing the Universal Declaration of Human Rights (1948), including Article 27 emphasizing the right to share in scientific advancement and its benefits; the instruments of the international human rights framework, including the International Convention on the Elimination of All Forms of Racial Discrimination (1965), the International Covenant on Civil and Political Rights (1966), the International Covenant on Economic, Social and Cultural Rights (1966), the United Nations Convention on the Elimination of All Forms of Discrimination against Women (1979), the United Nations Convention on the Rights of the Child (1989), and the United Nations Convention on the Rights of Persons with Disabilities (2006); the UNESCO Convention on the Protection and Promotion of the Diversity of Cultural Expressions (2005),

Noting the UNESCO Declaration on the Responsibilities of the Present Generations Towards Future Generations (1997); the United Nations Declaration on the Rights of Indigenous Peoples (2007); the Report of the United Nations Secretary-General on the Follow-up to the Second World Assembly on Ageing (A/66/173) of 2011, focusing on the situation of the human rights of older persons; the Report of the Special Representative of the United Nations Secretary-General on the issue of human rights and transnational corporations and other business enterprises (A/HRC/17/31) of 2011, outlining the "Guiding Principles on Business and Human Rights: Implementing United Nations 'Protect, Respect and Remedy' Framework"; the United Nations General Assembly resolution on the review of the World Summit on the Information Society (A/68/302); the Human Rights Council's resolution on "The right to privacy in the digital age" (A/HRC/RES/42/15) adopted on 26 September 2019; the Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (A/73/348); the UNESCO Recommendation on Science and Scientific Researchers (2017); the UNESCO Internet Universality Indicators (endorsed by UNESCO's International Programme for the Development of Communication in 2019), including the R.O.A.M. principles (endorsed by UNESCO's General Conference in 2015); the UNESCO Recommendation Concerning the Preservation of, and Access to, Documentary Heritage Including in Digital Form (2015); the Report of the United Nations Secretary-General's High-level Panel on Digital Cooperation on "The Age of Digital Interdependence" (2019), and the United Nations Secretary-General's Roadmap for Digital Cooperation (2020); the Universal Declaration on Bioethics and Human Rights (2005); the UNESCO Declaration on Ethical Principles in relation to Climate Change (2017); the United Nations Global Pulse initiative; and the outcomes and reports of the ITU's AI for Good Global Summits,

Noting also existing frameworks related to the ethics of AI of other intergovernmental organizations, such as the relevant human rights and other legal instruments adopted by the Council of Europe, and the work of its Ad Hoc Committee on AI (CAHAI); the work of the European Union related to AI, and of the European Commission's High-Level Expert Group on AI, including the Ethical Guidelines for Trustworthy AI; the work of OECD's first Group of Experts (AIGO) and its successor the OECD Network of Experts on AI (ONE AI), the OECD's Recommendation of the Council on AI and the OECD AI Policy Observatory (OECD.AI); the G20 AI Principles, drawn therefrom, and outlined in the G20 Ministerial Statement on Trade and Digital Economy; the G7's Charlevoix Common Vision for the Future of AI; the work of the African Union's Working Group on AI; and the work of the Arab League's Working Group on AI,

Emphasizing that specific attention must be paid to LMICs, including but not limited to LDCs, LLDCs and SIDS, as they have their own capacity but have been underrepresented in the AI ethics debate, which raises concerns about neglecting local knowledge, cultural and ethical pluralism, value systems and the demands of global fairness to deal with the positive and negative impacts of AI technologies,

Conscious of the many existing national policies and other frameworks related to the ethics and regulation of AI technologies,

Conscious as well of the many initiatives and frameworks related to the ethics of AI developed by the private sector, professional organizations, and non-governmental organizations, such as the IEEE's Global Initiative on Ethics of Autonomous and Intelligent Systems and its work on Ethically Aligned Design; the World Economic Forum's "Global Technology Governance: A Multistakeholder Approach"; the UNI Global Union's "Top 10 Principles for Ethical Artificial Intelligence"; the Montreal Declaration for a Responsible Development of AI; the Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems; the Harmonious Artificial Intelligence Principles (HAIP); and the Tenets of the Partnership on AI,

Convinced that AI technologies can bring important benefits, but that achieving them can also amplify tension around innovation debt, asymmetric access to knowledge, barriers of rights to information and gaps in capacity of creativity in developing cycles, human and institutional capacities, barriers to access to technological innovation, and a lack of adequate physical and digital infrastructure and regulatory frameworks regarding data,

Underlining that global cooperation and solidarity are needed to address the challenges that AI technologies bring in diversity and interconnectivity of cultures and ethical systems, to mitigate potential misuse, and to ensure that AI strategies and regulatory frameworks are not guided only by national and commercial interests and economic competition,

Taking fully into account that the rapid development of AI technologies challenges their ethical implementation and governance, because of the diversity of ethical orientations and cultures around the world, the lack of agility of the law in relation to technology and knowledge societies, and the risk that local and regional ethical standards and values be disrupted by AI technologies,

1. ***Adopts*** the present Recommendation on the Ethics of Artificial Intelligence;

2. ***Recommends*** that Member States apply the provisions of this Recommendation by taking appropriate steps, including whatever legislative or other measures may be required, in conformity with the constitutional practice and governing structures of each State, to give effect within their jurisdictions to the principles and norms of the Recommendation in conformity with international law, as well as constitutional practice;

3. ***Also recommends*** that Member States ensure assumption of responsibilities by all stakeholders, including private sector companies in AI technologies, and bring the Recommendation to the attention of the authorities, bodies, research and academic organizations, institutions and

organizations in public, private and civil society sectors involved in AI technologies, in order to guarantee that the development and use of AI technologies are guided by both sound scientific research as well as ethical analysis and evaluation.

I. SCOPE OF APPLICATION

1. This Recommendation addresses ethical issues related to AI. It approaches AI ethics as a systematic normative reflection, based on a holistic and evolving framework of interdependent values, principles and actions that can guide societies in dealing responsibly with the known and unknown impacts of AI technologies on human beings, societies, and the environment and ecosystems, and offers them a basis to accept or reject AI technologies. Rather than equating ethics to law, human rights, or a normative add-on to technologies, it considers ethics as a dynamic basis for the normative evaluation and guidance of AI technologies, referring to human dignity, well-being and the prevention of harm as a compass and rooted in the ethics of science and technology.

2. This Recommendation does not have the ambition to provide one single definition of AI, since such a definition would need to change over time, in accordance with technological developments. Rather, its ambition is to address those features of AI systems that are of central ethical relevance and on which there is large international consensus. Therefore, this Recommendation approaches AI systems as technological systems which have the capacity to process information in a way that resembles intelligent behaviour, and typically includes aspects of reasoning, learning, perception, prediction, planning or control. Three elements have a central place in this approach:

(a) AI systems are information-processing technologies that embody models and algorithms that produce a capacity to learn and to perform cognitive tasks leading to outcomes such as prediction and decision-making in real and virtual environments. AI systems are designed to operate with some aspects of autonomy by means of knowledge modelling and representation and by exploiting data and calculating correlations. AI systems may include several methods, such as but not limited to:

(i) machine learning, including deep learning and reinforcement learning,

(ii) machine reasoning, including planning, scheduling, knowledge representation and reasoning, search, and optimization, and

(iii) cyber-physical systems, including the Internet-of-Things, robotic systems, social robotics, and human-computer interfaces which involve control, perception, the processing of data collected by sensors, and the operation of actuators in the environment in which AI systems work.

(b) Ethical questions regarding AI systems pertain to all stages of the AI system life cycle, understood here to range from research, design, and development to deployment and use, including maintenance, operation, trade, financing, monitoring and evaluation, validation, end-of-use, disassembly, and termination. In addition, AI actors can be defined as any actor involved in at least one stage of the AI life cycle, and can refer both to natural and legal persons, such as researchers, programmers, engineers, data scientists, end-users, large technology companies, small and medium enterprises, start-ups, universities, public entities, among others.

(c) AI systems raise new types of ethical issues that include, but are not limited to, their impact on decision-making, employment and labour, social interaction, health care, education, media, freedom of expression, access to information, privacy, democracy, discrimination, and weaponization. Furthermore, new ethical challenges are created by the potential of AI algorithms to reproduce biases, for instance regarding gender, ethnicity, and age, and thus to exacerbate already existing forms of discrimination, identity prejudice and stereotyping. Some of these issues are related to the capacity of

AI systems to perform tasks which previously only living beings could do, and which were in some cases even limited to human beings only. These characteristics give AI systems a profound, new role in human practices and society, as well as in their relationship with the environment and ecosystems, creating a new context for children and young people to grow up in, develop an understanding of the world and themselves, critically understand media and information, and learn to make decisions. In the long term, AI systems could challenge human's special sense of experience and agency, raising additional concerns about human self-understanding, social, cultural and environmental interaction, autonomy, agency, worth and dignity.

3. This Recommendation pays specific attention to the broader ethical implications of AI systems in relation to the central domains of UNESCO: education, science, culture, and communication and information, as explored in the 2019 Preliminary Study on the Ethics of Artificial Intelligence by the UNESCO World Commission on Ethics of Scientific Knowledge and Technology (COMEST):

(a) Education, because living in digitalizing societies requires new educational practices, the need for ethical reflection, critical thinking, responsible design practices, and new skills, given the implications for the labour market and employability.

(b) Science, in the broadest sense and including all academic fields from the natural sciences and medical sciences to the social sciences and humanities, as AI technologies bring new research capacities, have implications for our concepts of scientific understanding and explanation, and create a new basis for decision-making.

(c) Cultural identity and diversity, as AI technologies can enrich cultural and creative industries, but can also lead to an increased concentration of supply of cultural content, data, markets, and income in the hands of only a few actors, with potential negative implications for the diversity and pluralism of languages, media, cultural expressions, participation and equality.

(d) Communication and information, as AI technologies play an increasingly important role in the processing, structuring and provision of information, and the issues of automated journalism and the algorithmic provision of news and moderation and curation of content on social media and search engines are just a few examples raising issues related to access to information, disinformation, misinformation, misunderstanding, the emergence of new forms of societal narratives, discrimination, freedom of expression, privacy, and media and information literacy, among others.

4. This Recommendation is addressed to States, both as AI actors and as responsible for developing legal and regulatory frameworks throughout the entire AI system life cycle, and for promoting business responsibility. It also provides ethical guidance to all AI actors, including the private sector, by providing a basis for an Ethical Impact Assessment of AI systems throughout their life cycle.

II. AIMS AND OBJECTIVES

5. This Recommendation aims to provide a basis to make AI systems work for the good of humanity, individuals, societies, and the environment and ecosystems; and to prevent harm.

6. In addition to the ethical frameworks regarding AI that have already been developed by various organizations all over the world, this Recommendation aims to bring a globally accepted normative instrument that does not only focus on the articulation of values and principles, but also on their practical realization, via concrete policy recommendations, with a strong emphasis on issues of gender equality and protection of the environment and ecosystems.

7. Because the complexity of the ethical issues surrounding AI necessitates the cooperation of multiple stakeholders across the various levels and sectors of international, regional and national communities, this Recommendation aims to enable stakeholders to take shared responsibility based on a global and intercultural dialogue.

8. The objectives of this Recommendation are:

(a) to provide a universal framework of values, principles and actions to guide States in the formulation of their legislation, policies or other instruments regarding AI;

(b) to guide the actions of individuals, groups, communities, institutions and private sector companies to ensure the embedding of ethics in all stages of the AI system life cycle;

(c) to promote respect for human dignity and gender equality, to safeguard the interests of present and future generations, and to protect human rights, fundamental freedoms, and the environment and ecosystems in all stages of the AI system life cycle;

(d) to foster multi-stakeholder, multidisciplinary and pluralistic dialogue about ethical issues relating to AI systems; and

(e) to promote equitable access to developments and knowledge in the field of AI and the sharing of benefits, with particular attention to the needs and contributions of LMICs, including LDCs, LLDCs and SIDS.

III. VALUES AND PRINCIPLES

9. The values and principles included below should be respected by all actors in the AI system life cycle, in the first place, and be promoted through amendments to existing and elaboration of new legislation, regulations and business guidelines. This must comply with international law as well as with international human rights law, principles and standards, and should be in line with social, political, environmental, educational, scientific and economic sustainability objectives.

10. Values play a powerful role as motivating ideals in shaping policy measures and legal norms. While the set of values outlined below thus inspires desirable behaviour and represents the foundations of principles, the principles unpack the values underlying them more concretely so that the values can be more easily operationalized in policy statements and actions.

11. While all the values and principles outlined below are desirable per se, in any practical context there are inevitable trade-offs among them, requiring complex choices to be made about contextual prioritization, without compromising other principles or values in the process. Trade-offs should take account of concerns related to proportionality and legitimate purpose. To navigate such scenarios judiciously will typically require engagement with a broad range of appropriate stakeholders guided by international human rights law, standards and principles, making use of social dialogue, as well as ethical deliberation, due diligence, and impact assessment.

12. The trustworthiness and integrity of the life cycle of AI systems, if achieved, work for the good of humanity, individuals, societies, and the environment and ecosystems, and embody the values and principles set out in this Recommendation. People should have good reason to trust that AI systems bring shared benefits, while adequate measures are taken to mitigate risks. An essential requirement for trustworthiness is that, throughout their life cycle, AI systems are subject to monitoring by governments, private sector companies, independent civil society and other stakeholders. As trustworthiness is an outcome of the operationalization of the principles in this document, the policy actions proposed in this Recommendation are all directed at promoting trustworthiness in all stages of the AI life cycle.

III.1. VALUES

Respect, protection and promotion of human dignity, human rights and fundamental freedoms

13. The dignity of every human person constitutes a foundation for the indivisible system of human rights and fundamental freedoms and is essential throughout the life cycle of AI systems. Human dignity relates to the recognition of the intrinsic worth of each individual human being and thus dignity is not tied to sex, gender, language, religion, political or other opinion, national, ethnic, indigenous or social origin, sexual orientation and gender identity, property, birth, disability, age or other status.

14. No human being should be harmed physically, economically, socially, politically, or mentally during any phase of the life cycle of AI systems. Throughout the life cycle of AI systems the quality of life of every human being should be enhanced, while the definition of “quality of life” should be left open to individuals or groups, as long as there is no violation or abuse of human rights, or the dignity of humans in terms of this definition.

15. Persons may interact with AI systems throughout their life cycle and receive assistance from them such as care for vulnerable people, including but not limited to children, older persons, persons with disabilities or the ill. Within such interactions, persons should never be objectified, nor should their dignity be undermined, or human rights violated or abused.

16. Human rights and fundamental freedoms must be respected, protected, and promoted throughout the life cycle of AI systems. Governments, private sector, civil society, international organizations, technical communities, and academia must respect human rights instruments and frameworks in their interventions in the processes surrounding the life cycle of AI systems. New technologies need to provide new means to advocate, defend and exercise human rights and not to infringe them.

Environment and ecosystem flourishing

17. Environmental and ecosystem flourishing should be recognized and promoted through the life cycle of AI systems. Furthermore, environment and ecosystems are the existential necessity for humanity and other living beings to be able to enjoy the benefits of advances in AI.

18. All actors involved in the life cycle of AI systems must follow relevant international law and domestic legislation, standards and practices, such as precaution, designed for environmental and ecosystem protection and restoration, and sustainable development. They should reduce the environmental impact of AI systems, including but not limited to, its carbon footprint, to ensure the minimization of climate change and environmental risk factors, and prevent the unsustainable exploitation, use and transformation of natural resources contributing to the deterioration of the environment and the degradation of ecosystems.

Ensuring diversity and inclusiveness

19. Respect, protection and promotion of diversity and inclusiveness should be ensured throughout the life cycle of AI systems, at a minimum consistent with international human rights law, standards and principles, as well as demographic, cultural, gender and social diversity and inclusiveness. This may be done by promoting active participation of all individuals or groups based on sex, gender, language, religion, political or other opinion, national, ethnic, indigenous or social origin, sexual orientation and gender identity, property, birth, disability, age or other status, in the life cycle of AI systems. Any homogenizing tendency should be monitored and addressed.

20. The scope of lifestyle choices, beliefs, opinions, expressions or personal experiences, including the optional use of AI systems and the co-design of these architectures should not be restricted in any way during any phase of the life cycle of AI systems.

21. Furthermore, efforts should be made to overcome, and never exploit, the lack of necessary technological infrastructure, education and skills, as well as legal frameworks, in some communities, and particularly in LMICs, LDCs, LLDCs and SIDS.

Living in harmony and peace

22. AI actors should play an enabling role for harmonious and peaceful life, which is to ensure an interconnected future ensuring the benefit of all. The value of living in harmony and peace points to the potential of AI systems to contribute throughout their life cycle to the interconnectedness of all living creatures with each other and with the natural environment.

23. The notion of humans being interconnected is based on the knowledge that every human belongs to a greater whole, which is diminished when others are diminished in any way. Living in harmony and peace requires an organic, immediate, uncalculated bond of solidarity, characterized by a permanent search for non-conflictual, peaceful relations, tending towards consensus with others and harmony with the natural environment in the broadest sense of the term.

24. This value demands that peace should be promoted throughout the life cycle of AI systems, in so far as the processes of the life cycle of AI systems should not segregate, objectify, or undermine the safety of human beings, divide and turn individuals and groups against each other, or threaten the harmonious coexistence between humans, non-humans, and the natural environment, as this would negatively impact on humankind as a collective.

III.2. PRINCIPLES

Proportionality and do no harm

25. It should be recognized that AI technologies do not necessarily, per se, ensure human and environmental and ecosystem flourishing. Furthermore, none of the processes related to the AI system life cycle shall exceed what is necessary to achieve legitimate aims or objectives and should be appropriate to the context. In the event of possible occurrence of any harm to human beings or the environment and ecosystems, the implementation of procedures for risk assessment and the adoption of measures in order to preclude the occurrence of such harm should be ensured.

26. The choice of an AI method should be justified in the following ways: (a) The AI method chosen should be desirable and proportional to achieve a given legitimate aim; (b) The AI method chosen should not have a negative infringement on the foundational values captured in this document; (c) The AI method should be appropriate to the context and should be based on rigorous scientific foundations. In scenarios that involve life and death decisions, final human determination should apply.

Safety and security

27. Unwanted harms (safety risks) and vulnerabilities to attacks (security risks) should be avoided throughout the life cycle of AI systems to ensure human and environmental and ecosystem safety and security. Safe and secure AI will be enabled by the development of sustainable, privacy-protective data access frameworks that foster better training of AI models utilizing quality data.

Fairness and non-discrimination

28. AI actors should promote social justice, by respecting fairness. Fairness implies sharing benefits of AI technologies at local, national and international levels, while taking into consideration the specific needs of different age groups, cultural systems, different language groups, persons with disabilities, girls and women, and disadvantaged, marginalized and vulnerable populations. At the local level, it is a matter of working to give communities access to AI systems in the languages of their choice and respecting different cultures. At the national level, governments are obliged to demonstrate equity between rural and urban areas, and among all persons without distinction as to

sex, gender, language, religion, political or other opinion, national, ethnic, indigenous or social origin, sexual orientation and gender identity, property, birth, disability, age or other status, in terms of access to and participation in the AI system life cycle. At the international level, the most technologically advanced countries have an obligation of solidarity with the least advanced to ensure that the benefits of AI technologies are shared such that access to and participation in the AI system life cycle for the latter contributes to a fairer world order with regard to information, communication, culture, education, research, and socio-economic and political stability.

29. AI actors should make all efforts to minimize and avoid reinforcing or perpetuating inappropriate socio-technical biases based on identity prejudice, throughout the life cycle of the AI system to ensure fairness of such systems. There should be a possibility to have a remedy against unfair algorithmic determination and discrimination.

30. Furthermore, discrimination, digital and knowledge divides, and global inequalities need to be addressed throughout an AI system life cycle, including in terms of access to technology, data, connectivity, knowledge and skills, and participation of the affected communities as part of the design phase, such that every person is treated equitably.

Sustainability

31. The development of sustainable societies relies on the achievement of a complex set of objectives on a continuum of social, cultural, economic and environmental dimensions. The advent of AI technologies can either benefit sustainability objectives or hinder their realization, depending on how they are applied across countries with varying levels of development. The continuous assessment of the social, cultural, economic and environmental impact of AI technologies should therefore be carried out with full cognizance of the implications of AI technologies for sustainability as a set of constantly evolving goals across a range of dimensions, such as currently identified in the United Nations Sustainable Development Goals (SDGs).

Privacy

32. Privacy, a right essential to the protection of human dignity, human autonomy and human agency, must be respected, protected and promoted throughout the life cycle of AI systems both at the personal and collective level. It is crucial that data for AI is being collected, used, shared, archived and deleted in ways that are consistent with the values and principles set forth in this Recommendation.

33. Adequate data protection frameworks and governance mechanisms should be established by regulatory agencies, at national or supranational level, protected by judicial systems, and ensured throughout the life cycle of AI systems. This protection framework and mechanisms concern the collection, control over, and use of data and exercise of their rights by data subjects and of the right for individuals to have personal data erased, ensuring a legitimate aim and a valid legal basis for the processing of personal data as well as for the personalization, and de- and re-personalization of data, transparency, appropriate safeguards for sensitive data, and effective independent oversight.

34. Algorithmic systems require thorough privacy impact assessments which also include societal and ethical considerations of their use and an innovative use of the privacy by design approach.

Human oversight and determination

35. It must always be possible to attribute ethical and legal responsibility for any stage of the life cycle of AI systems to physical persons or to existing legal entities. Human oversight refers thus not only to individual human oversight, but to public oversight, as appropriate.

36. It may be the case that sometimes humans would have to rely on AI systems for reasons of efficacy, but the decision to cede control in limited contexts remains that of humans, as humans can

resort to AI systems in decision-making and acting, but an AI system can never replace ultimate human responsibility and accountability.

Transparency and explainability

37. The transparency of AI systems is often a crucial precondition to ensure that fundamental human rights and ethical principles are respected, protected and promoted. Transparency is necessary for relevant national and international liability legislation to work effectively.

38. While efforts need to be made to increase transparency and explainability of AI systems throughout their life cycle to support democratic governance, the level of transparency and explainability should always be appropriate to the context, as some trade-offs exist between transparency and explainability and other principles such as safety and security. People have the right to be aware when a decision is being made on the basis of AI algorithms, and in those circumstances require or request explanatory information from private sector companies or public sector institutions.

39. From a socio-technical lens, greater transparency contributes to more peaceful, just and inclusive societies. It allows for public scrutiny that can decrease corruption and discrimination, and can also help detect and prevent negative impacts on human rights. Transparency may contribute to trust from humans for AI systems. Specific to the AI system, transparency can enable people to understand how each stage of an AI system is put in place, appropriate to the context and sensitivity of the AI system. It may also include insight into factors that impact a specific prediction or decision, and whether or not appropriate assurances (such as safety or fairness measures) are in place. In cases where serious adverse human rights impacts are foreseen, transparency may also require the sharing of specific code or datasets.

40. Explainability refers to making intelligible and providing insight into the outcome of AI systems. The explainability of AI systems also refers to the understandability of the input, output and behaviour of each algorithmic building block and how it contributes to the outcome of the systems. Thus, explainability is closely related to transparency, as outcomes and sub-processes leading to outcomes should be understandable and traceable, appropriate to the use context.

41. Transparency and explainability relate closely to adequate responsibility and accountability measures, as well as to the trustworthiness of AI systems.

Responsibility and accountability

42. AI actors should respect, protect and promote human rights and promote the protection of the environment and ecosystems, assuming ethical and legal responsibility in accordance with extant national and international law, in particular international human rights law, principles and standards, and ethical guidance throughout the life cycle of AI systems. The ethical responsibility and liability for the decisions and actions based in any way on an AI system should always ultimately be attributable to AI actors.

43. Appropriate oversight, impact assessment, and due diligence mechanisms should be developed to ensure accountability for AI systems and their impact throughout their life cycle. Both technical and institutional designs should ensure auditability and traceability of (the working of) AI systems in particular to address any conflicts with human rights and threats to environmental and ecosystem well-being.

Awareness and literacy

44. Public awareness and understanding of AI technologies and the value of data should be promoted through open and accessible education, civic engagement, digital skills and AI ethics training, media and information literacy and training led jointly by governments, intergovernmental organizations, civil society, academia, the media, community leaders and the private sector, and

considering the existing linguistic, social and cultural diversity, to ensure effective public participation so that all members of society can take informed decisions about their use of AI systems and be protected from undue influence.

45. Learning about the impact of AI systems should include learning about, through and for human rights, meaning that the approach and understanding of AI systems should be grounded by their impact on human rights and access to rights.

Multi-stakeholder and adaptive governance and collaboration

46. International law and sovereignty should be respected in the use of data. Data sovereignty means that States, complying with international law, regulate the data generated within or passing through their territories, and take measures towards effective regulation of data based on respect for the right to privacy and other human rights.

47. Participation of different stakeholders throughout the AI system life cycle is necessary for inclusive AI governance, sharing of benefits of AI, and fair technological advancement and its contribution to development goals. Stakeholders include but are not limited to governments, intergovernmental organizations, the technical community, civil society, researchers and academia, media, education, policy-makers, private sector companies, human rights institutions and equality bodies, anti-discrimination monitoring bodies, and groups for youth and children. The adoption of open standards and interoperability to facilitate collaboration must be in place. Measures must be adopted to take into account shifts in technologies, the emergence of new groups of stakeholders, and to allow for meaningful intervention by marginalized groups, communities and individuals.

IV. AREAS OF POLICY ACTION

48. The policy actions described in the following policy areas operationalize the values and principles set out in this Recommendation. The main action is for Member States to put in place policy frameworks or mechanisms and to ensure that other stakeholders, such as private sector companies, academic and research institutions, and civil society, adhere to them by, among other actions, assisting all stakeholders to develop ethical impact assessment and due diligence tools. The process for developing such policies or mechanisms should be inclusive of all stakeholders and should take into account the circumstances and priorities of each Member State. UNESCO can be a partner and support Member States in the development as well as monitoring and evaluation of policy mechanisms.

49. UNESCO recognizes that Member States will be at different stages of readiness to implement this Recommendation, in terms of scientific, technological, economic, educational, legal, regulatory, infrastructural, societal, cultural and other dimensions. It is noted that “readiness” here is a dynamic status. In order to enable the effective implementation of this Recommendation, UNESCO will therefore: (1) develop a readiness assessment methodology to assist Member States in identifying their status at specific moments of their readiness trajectory along a continuum of dimensions; and (2) ensure support for Member States in terms of developing a globally accepted methodology for Ethical Impact Assessment (EIA) of AI technologies, sharing of best practices, assessment guidelines and other mechanisms and analytical work.

POLICY AREA 1: ETHICAL IMPACT ASSESSMENT

50. Member States should introduce impact assessments to identify and assess benefits, concerns and risks of AI systems, as well as risk prevention, mitigation and monitoring measures. The ethical impact assessment should identify impacts on human rights, in particular but not limited to the rights of vulnerable groups, labour rights, the environment and ecosystems, and ethical and social implications in line with the principles set forth in this Recommendation.

51. Member States and private sector companies should develop due diligence and oversight mechanisms to identify, prevent, mitigate and account for how they address the impact of AI systems on human rights, rule of law and inclusive societies. Member States should also be able to assess the socio-economic impact of AI systems on poverty and ensure that the gap between people living in wealth and poverty, as well as the digital divide among and within countries are not increased with the massive adoption of AI technologies at present and in the future. In order to do this, enforceable transparency protocols should be implemented, corresponding to the right of access to information, including information of public interest held by private entities.

52. Member States and private sector companies should implement proper measures to monitor all phases of an AI system life cycle, including the behaviour of algorithms used for decision-making, the data, as well as AI actors involved in the process, especially in public services and where direct end-user interaction is needed.

53. Governments should adopt a regulatory framework that sets out a procedure, particularly for public authorities, to carry out ethical impact assessments on AI systems to predict consequences, mitigate risks, avoid harmful consequences, facilitate citizen participation and address societal challenges. The assessment should also establish appropriate oversight mechanisms, including auditability, traceability and explainability which enable the assessment of algorithms, data and design processes, as well as include external review of AI systems. Ethical impact assessments carried out by public authorities should be transparent and open to the public. Such assessments should also be multidisciplinary, multi-stakeholder, multicultural, pluralistic and inclusive. Member States are encouraged to put in place mechanisms and tools, for example regulatory sandboxes or testing centres, which would enable impact monitoring and assessment in a multidisciplinary and multi-stakeholder fashion. The public authorities should be required to monitor the AI systems implemented and/or deployed by those authorities by introducing appropriate mechanisms and tools.

54. Member States should establish monitoring and evaluation mechanisms for initiatives and policies related to AI ethics. Possible mechanisms include: a repository covering human rights-compliant and ethical development of AI systems; a lessons sharing mechanism for Member States to seek feedback from other Member States on their policies and initiatives; a guide for all AI actors to assess their adherence to policy recommendations mentioned in this document; and follow-up tools. International human rights law, standards and principles should form part of the ethical aspects of AI system assessments.

POLICY AREA 2: ETHICAL GOVERNANCE AND STEWARDSHIP

55. Member States should ensure that any AI governance mechanism is inclusive, transparent, multidisciplinary, multilateral (this includes the possibility of mitigation and redress of harm across borders), and multi-stakeholder. Governance should include aspects of anticipation, protection, monitoring of impact, enforcement and redressal.

56. Member States should ensure that harms caused to users through AI systems are investigated and redressed, by enacting strong enforcement mechanisms and remedial actions, to make certain that human rights and the rule of law are respected in the digital world as it is in the physical world. Such mechanisms and actions should include remediation mechanisms provided by private sector companies. The auditability and traceability of AI systems should be promoted to this end. In addition, Member States should strengthen their institutional capacities to deliver on this duty of care and should collaborate with researchers and other stakeholders to investigate, prevent and mitigate any potentially malicious uses of AI systems.

57. Member States are encouraged to consider forms of soft governance such as a certification mechanism for AI systems and the mutual recognition of their certification, according to the sensitivity of the application domain and expected impact on human lives, the environment and ecosystems, and other ethical considerations set forth in this Recommendation. Such a mechanism might include different levels of audit of systems, data, and adherence to ethical guidelines, and should be

validated by authorized parties in each country. At the same time, such a mechanism must not hinder innovation or disadvantage small and medium enterprises or start-ups by requiring large amounts of paperwork. These mechanisms would also include a regular monitoring component to ensure system robustness and continued integrity and adherence to ethical guidelines over the entire life cycle of the AI system, requiring re-certification if necessary.

58. Government and public authorities should be required to carry out self-assessment of existing and proposed AI systems, which in particular, should include the assessment whether the adoption of AI is appropriate and, if so, should include further assessment to determine what the appropriate method is, as well as assessment as to whether such adoption transgresses any human rights law, standards and principles.

59. Member States should encourage public entities, private sector companies and civil society organizations to involve different stakeholders in their AI governance and to consider adding the role of an independent AI Ethics Officer or some other mechanism to oversee ethical impact assessment, auditing and continuous monitoring efforts and ensure ethical guidance of AI systems. Member States, private sector companies and civil society organizations, with the support of UNESCO, are encouraged to create a network of independent AI Ethics Officers to give support to this process at national, regional and international levels.

60. Member States should foster the development of, and access to, a digital ecosystem for ethical development of AI systems at the national level, while encouraging international collaboration. Such an ecosystem includes in particular digital technologies and infrastructure, and mechanisms for sharing AI knowledge, as appropriate. In this regard, Member States should consider reviewing their policies and regulatory frameworks, including on access to information and open government to reflect AI-specific requirements and promoting mechanisms, such as open repositories for publicly-funded or publicly-held data and source code and data trusts, to support the safe, fair, legal and ethical sharing of data, among others.

61. Member States should establish mechanisms, in collaboration with international organizations, transnational corporations, academic institutions and civil society, to ensure the active participation of all Member States, especially LMICs, in particular LDCs, LLDCs and SIDS, in international discussions concerning AI governance. This can be through the provision of funds, ensuring equal regional participation, or any other mechanisms. Furthermore, in order to ensure the inclusiveness of AI fora, Member States should facilitate the travel of AI actors in and out of their territory, especially from LMICs, in particular LDCs, LLDCs and SIDS, for the purpose of participating in these fora.

62. Amendments to existing or elaboration of new national legislation addressing AI systems must comply with international human rights law and promote human rights and fundamental freedoms throughout the AI system life cycle. Promotion thereof should also take the form of governance initiatives, good exemplars of collaborative practices regarding AI systems, and national and international technical and methodological guidelines as AI technologies advance. Diverse sectors, including the private sector, in their practices regarding AI systems must respect, protect and promote human rights and fundamental freedoms using existing and new instruments in combination with this Recommendation.

63. Member States should provide mechanisms for human rights and for social and economic impact of AI monitoring and oversight, and other governance mechanisms such as independent data protection authorities, sectoral oversight, public bodies for the oversight of acquisition of AI systems for human rights sensitive use cases, such as criminal justice, law enforcement, welfare, employment, health care, among others, and independent judiciary systems.

64. Member States should ensure that governments and multilateral organizations play a leading role in guaranteeing the safety and security of AI systems. Specifically, Member States, international organizations and other relevant bodies should develop international standards that describe measurable, testable levels of safety and transparency, so that systems can be objectively assessed

and levels of compliance determined. Furthermore, Member States should continuously support strategic research on potential safety and security risks of AI technologies and should encourage research into transparency and explainability by putting additional funding into those areas for different domains and at different levels, such as technical and natural language.

65. Member States should implement policies to ensure that the actions of AI actors are consistent with international human rights law, standards and principles throughout the life cycle of AI systems, while demonstrating awareness and respect for the current cultural and social diversities including local customs and religious traditions.

66. Member States should put in place mechanisms to require AI actors to disclose and combat any kind of stereotyping in the outcomes of AI systems and data, whether by design or by negligence, and to ensure that training data sets for AI systems do not foster cultural, economic or social inequalities, prejudice, the spreading of non-reliable information or the dissemination of anti-democratic ideas. Particular attention should be given to regions where the data are scarce.

67. Member States should implement policies to promote and increase diversity in AI development teams and training datasets, and to ensure equal access to AI technologies and their benefits, particularly for marginalized groups, both from rural and urban zones.

68. Member States should develop, review and adapt, as appropriate, regulatory and legal frameworks to achieve accountability and responsibility for the content and outcomes of AI systems at the different phases of their life cycle. Member States should introduce liability frameworks or clarify the interpretation of existing frameworks to ensure the attribution of accountability for the outcomes and behaviour of AI systems. Furthermore, when developing regulatory frameworks, Member States should, in particular, take into account that ultimate responsibility and accountability must always lie with natural or legal persons and that AI systems should not be given legal personality themselves. To ensure this, such regulatory frameworks should be consistent with the principle of human oversight and establish a comprehensive approach focused on the actors and the technological processes involved across the different stages of the AI systems life cycle.

69. Member States should enhance the capacity of the judiciary to make decisions related to AI systems as per the rule of law and in line with international standards, including in the use of AI systems in their deliberations, while ensuring that the principle of human oversight is upheld.

70. In order to establish norms where these do not exist, or to adapt existing legal frameworks, Member States should involve all AI actors (including, but not limited to, researchers, representatives of civil society and law enforcement, insurers, investors, manufacturers, engineers, lawyers, and users). The norms can mature into best practices, laws and regulations. Member States are further encouraged to use mechanisms such as policy prototypes and regulatory sandboxes to accelerate the development of laws, regulations and policies in line with the rapid development of new technologies and ensure that laws and regulations can be tested in a safe environment before being officially adopted. Member States should support local governments in the development of local policies, regulations, and laws in line with national and international legal frameworks.

71. Member States should set clear requirements for AI system transparency and explainability so as to help ensure the trustworthiness of the full AI system life cycle. Such requirements should involve the design and implementation of impact mechanisms that take into consideration the nature of application domain (Is this a high-risk domain such as law enforcement, security, education, recruitment and health care?), intended use (What are the risks in terms of transgression of safety and human rights?), target audience (Who is requesting the information) and feasibility (Is the algorithm explainable or not and what are the trade-offs between accuracy and explainability?) of each particular AI system.

POLICY AREA 3: DATA POLICY

72. Member States should work to develop data governance strategies that ensure the continual evaluation of the quality of training data for AI systems including the adequacy of the data collection and selection processes, proper security and data protection measures, as well as feedback mechanisms to learn from mistakes and share best practices among all AI actors. Striking a balance between the collection of metadata and users' privacy should be an upfront goal for such a strategy.

73. Member States should put in place appropriate safeguards to recognize and protect individuals' fundamental right to privacy, including through the adoption or the enforcement of legislative frameworks that provide appropriate protection, compliant with international law. Member States should strongly encourage all AI actors, including private sector companies, to follow existing international standards and in particular to carry out privacy impact assessments, as part of ethical impact assessments, which take into account the wider socio-economic impact of the intended data processing and to apply privacy by design in their systems. Privacy should be respected, protected and promoted throughout the life cycle of AI systems.

74. Member States should ensure that individuals retain rights over their personal data and are protected by a framework which notably foresees transparency, appropriate safeguards for the processing of sensitive data, the highest level of data security, effective and meaningful accountability schemes and mechanisms, the full enjoyment of data subjects' rights, in particular the right to access and the right to erasure of their personal data in AI systems, an appropriate level of protection while data are being used for commercial purposes such as enabling micro-targeted advertising, transferred cross-border, and an effective independent oversight as part of a data governance mechanism which respects data sovereignty and balances this with the benefits of a free flow of information internationally, including access to data.

75. Member States should establish their data policies or equivalent frameworks, or reinforce existing ones, to ensure increased security for personal data and sensitive data, which if disclosed, may cause exceptional damage, injury or hardship to a person. Examples include data relating to offences, criminal proceedings and convictions, and related security measures; biometric and genetic data; personal data relating to ethnic or social origin, political opinions, trade union membership, religious and other beliefs, health and sexual life.

76. Member States should use AI systems to improve access to information and knowledge, including of their data holdings, and address gaps in access to the AI system life cycle. This can include support to researchers and developers to enhance freedom of expression and access to information, and increased proactive disclosure of official data and information. Member States should also promote open data, including through developing open repositories for publicly-funded or publicly-held data and source code.

77. Member States should ensure the overall quality and robustness of the dataset for AI, and exercise vigilance in overseeing their collection and use. This could, if possible and feasible, include investing in the creation of gold standard datasets, including open and trustworthy datasets, which are diverse, constructed on a valid legal basis, including consent of data subjects, when required by law. Standards for annotating datasets should be encouraged, so it can easily be determined how a dataset is gathered and what properties it has.

78. Member States, as also suggested in the report of the UNSG's High-level Panel on Digital Cooperation, with the support of the United Nations and UNESCO, should adopt a Digital Commons approach to data where appropriate, increase interoperability of tools and datasets and interfaces of systems hosting data, and encourage private sector companies to share the data they collect as appropriate for research or public benefits. They should also promote public and private efforts to create collaborative platforms to share quality data in trusted and secured data spaces.

POLICY AREA 4: DEVELOPMENT AND INTERNATIONAL COOPERATION

79. Member States and transnational corporations should prioritize AI ethics by including discussions of AI-related ethical issues into relevant international, intergovernmental and multi-stakeholder fora.

80. Member States should ensure that the use of AI in areas of development such as health care, agriculture/food supply, education, media, culture, environment, water management, infrastructure management, economic planning and growth, and others, adheres to the values and principles set forth in this Recommendation.

81. Member States should work through international organizations to provide platforms for international cooperation on AI for development, including by contributing expertise, funding, data, domain knowledge, infrastructure, and facilitating collaboration between technical and business experts to tackle challenging development problems, especially for LMICs, in particular LDCs, LLDCs and SIDS.

82. Member States should work to promote international collaboration on AI research and innovation, including research and innovation centres and networks that promote greater participation and leadership of researchers from LMICs and other regions, including LDCs, LLDCs and SIDS.

83. Member States should promote AI ethics research by international organizations and research institutions, as well as transnational corporations, that can be a basis for the ethical use of AI systems by public and private entities, including research into the applicability of specific ethical frameworks in specific cultures and contexts, and the possibilities to match these frameworks to technologically feasible solutions.

84. Member States should encourage international cooperation and collaboration in the field of AI to bridge geo-technological lines. Technological exchanges/consultations should take place between Member States and their populations, between the public and private sectors, and between and among Member States in the Global North and Global South.

85. Member States should develop and implement an international legal framework to encourage international cooperation between States and other stakeholders paying special attention to the situation of LMICs, in particular LDCs, LLDCs and SIDS.

POLICY AREA 5: ENVIRONMENT AND ECOSYSTEMS

86. Member States should assess the direct and indirect environmental impact throughout the AI system life cycle, including but not limited to, its carbon footprint, energy consumption, and the environmental impact of raw material extraction for supporting the manufacturing of AI technologies. They should ensure compliance of all AI actors with environmental law, policies, and practices.

87. Member States should introduce incentives, when needed and appropriate, to ensure the development and adoption of rights-based and ethical AI-powered solutions for disaster risk resilience; the monitoring, protection and regeneration of the environment and ecosystems; and the preservation of the planet. These AI systems should involve the participation of local and indigenous communities throughout their life cycle and should support circular economy type approaches and sustainable consumption and production patterns. Some examples include using AI systems, when needed and appropriate, to:

(a) Support the protection, monitoring, and management of natural resources.

(b) Support the prevention, control, and management of climate-related problems.

(c) Support a more efficient and sustainable food ecosystem.

(d) Support the acceleration of access to and mass adoption of sustainable energy.

(e) Enable and promote the mainstreaming of sustainable infrastructure, sustainable business models, and sustainable finance for sustainable development.

(f) Detect pollutants or predict levels of pollution and thus help relevant stakeholders identify, plan and put in place targeted interventions to prevent and reduce pollution and exposure.

88. When choosing AI methods, given the data-intensive or resource-intensive character of some of them and the respective impact on the environment, Member States should ensure that AI actors, in line with the principle of proportionality, favour data, energy and resource-efficient AI methods. Requirements should be developed to ensure that appropriate evidence is available showing that an AI application will have the intended effect, or that safeguards accompanying an AI application can support the justification.

POLICY AREA 6: GENDER

89. Member States should ensure that digital technologies and artificial intelligence fully contribute to achieve gender equality; and that the rights and fundamental freedoms of girls and women, including their safety and integrity are not violated at any stage of the AI system life cycle. Moreover, Ethical Impact Assessments should include a transversal gender perspective.

90. Member States should have dedicated funds from the public budgets linked to financing gender-related schemes, ensure that national digital policies include a gender action plan, and develop specific policies, e.g. on labour education, targeted at supporting girls and women to make sure girls and women are not left out of the digital economy powered by AI. Special investment in providing targeted programmes and gender-specific language, to increase the opportunities of participation of girls and women in science, technology, engineering, and mathematics (STEM), including information and communication technologies (ICT) disciplines, preparedness, employability, career development and professional growth of girls and women should be considered and implemented.

91. Member States should ensure that the potential of AI systems to improve gender equality is realized. They should guarantee that these technologies do not contribute to exacerbating the already wide gender gaps existing in several fields in the analogue world. This includes the gender wage gap; the representation in certain professions and activities gap; the lack of representation at the top management positions, boards of directors, or research teams in the AI field; the education gap; digital/AI access, adoption, usage and affordability gap; the unequal distribution of unpaid work and of the caring responsibilities in our societies.

92. Member States should ensure that gender stereotyping, and discriminatory biases are not translated into the AI systems. Efforts are necessary to avoid the compounding negative effect of technological divides in achieving gender equality and avoiding violence against girls and women, and all other types of gender identities.

93. Member States should encourage female entrepreneurship, participation and engagement in all stages of an AI system life cycle by offering and promoting economic, regulatory incentives, among other incentives and support schemes, as well as policies that aim at a balanced gender participation in AI research in academia, gender representation on digital/AI companies top management positions, board of directors, or research teams. Governments should ensure public funds (on innovation, research and technologies) are channelled to inclusive programmes and companies, with clear gender representation, and that private funds are encouraged through affirmative action principles. Moreover, policies on harassment-free environments should be developed and enforced together with the encouragement of the transfer of best practices on how to promote diversity throughout the AI system life cycle.

94. UNESCO can help form a repository of best practices for incentivizing the participation of women and under-represented groups on all stages of the AI life cycle.

POLICY AREA 7: CULTURE

95. Member States are encouraged to incorporate AI systems where appropriate in the preservation, enrichment, understanding, promotion and accessibility of tangible, documentary and intangible cultural heritage, including endangered languages as well as indigenous languages and knowledge, for example by introducing or updating educational programmes related to the application of AI systems in these areas where appropriate and ensuring a participatory approach, targeted at institutions and the public.

96. Member States are encouraged to examine and address the cultural impact of AI systems, especially Natural Language Processing applications such as automated translation and voice assistants on the nuances of human language and expression. Such assessments should provide input for the design and implementation of strategies that maximize the benefits from these systems by bridging cultural gaps and increasing human understanding, as well as negative implications such as the reduction of use, which could lead to the disappearance of endangered languages, local dialects, and tonal and cultural variations associated with human language and expression.

97. Member States should promote AI education and digital training for artists and creative professionals to assess the suitability of AI technologies for use in their profession as AI technologies are being used to create, produce, distribute and broadcast a variety of cultural goods and services, bearing in mind the importance of preserving cultural heritage, diversity and artistic freedom.

98. Member States should promote awareness and evaluation of AI tools among local cultural industries and small and medium enterprises working in the field of culture, to avoid the risk of concentration in the cultural market.

99. Member States should engage large technology companies and other stakeholders to promote a diverse supply and plural access to cultural expressions, and in particular to ensure that algorithmic recommendation enhances the visibility and discoverability of local content.

100. Member States should foster new research at the intersection between AI and intellectual property, for example to determine who are the rights-holders of the works created by means of AI technologies among the different stakeholders throughout the AI life cycle.

101. Member States should encourage museums, galleries, libraries and archives at the national level to develop and use AI systems to highlight their collections, strengthen their databases and grant access to them for their users.

POLICY AREA 8: EDUCATION AND RESEARCH

102. Member States should work with international organizations, private and non-governmental entities to provide adequate AI literacy education to the public in all countries in order to empower people and reduce the digital divide and digital access inequalities resulting from the wide adoption of AI systems.

103. Member States should promote the acquisition of “prerequisite skills” for AI education, such as basic literacy, numeracy, coding and digital skills, and media and information literacy, as well as critical thinking, teamwork, communication, socio-emotional, and AI ethics skills, especially in countries where there are notable gaps in the education of these skills.

104. Member States should promote general awareness programmes about AI developments, including on the opportunities and challenges brought about by AI technologies. These programmes should be accessible to non-technical as well as technical groups.

105. Member States should encourage research initiatives on the responsible use of AI technologies in teaching, teacher training and e-learning among other topics, in a way that enhances opportunities and mitigates the challenges and risks involved in this area. The initiatives should be accompanied by an adequate assessment of the quality of education and of impact on students and teachers of the use of AI technologies. Member States should also ensure that AI technologies empower students and teachers and enhance their experience, bearing in mind that emotional and social aspects and the value of traditional forms of education are vital in the teacher-student and student-student relationships, and should be considered when discussing the adoption of AI technologies in education.

106. Member States should promote the participation of girls and women, diverse ethnicities and cultures, and persons with disabilities, in AI education programmes at all levels, as well as the monitoring and sharing of best practices in this regard with other Member States.

107. Member States should develop, in accordance with their national education programmes and traditions, AI ethics curricula for all levels, and promote cross-collaboration between AI technical skills education and humanistic, ethical and social aspects of AI education. Online courses and digital resources of AI ethics education should be developed in local languages, especially in accessible formats for persons with disabilities.

108. Member States should promote AI ethics research either through investing in such research or by creating incentives for the public and private sectors to invest in this area.

109. Member States should ensure that AI researchers are trained in research ethics and require them to include ethical considerations in their designs, products and publications, especially in the analyses of the datasets they use, how they are annotated and the quality and the scope of the results.

110. Member States should encourage private sector companies to facilitate the access of scientific community to their data for research, especially in LMICs, in particular LDCs, LLDCs and SIDS. This access should not be at the expense of privacy.

111. Member States should promote gender diversity in AI research in academia and industry by offering incentives to girls and women to enter the field, putting in place mechanisms to fight gender stereotyping and harassment within the AI research community, and encouraging academic and private entities to share best practices on how to enhance gender diversity.

112. To ensure a critical evaluation of AI research and proper monitoring of potential misuses or adverse effects, Member States should ensure that any future developments with regards to AI technologies should be based on rigorous scientific research, and promote interdisciplinary AI research by including disciplines other than science, technology, engineering, and mathematics (STEM), such as cultural studies, education, ethics, international relations, law, linguistics, philosophy, and political science.

113. Recognizing that AI technologies present great opportunities to help advance scientific knowledge and practice, especially in traditionally model-driven disciplines, Member States should encourage scientific communities to be aware of the benefits, limits and risks of their use; this includes attempting to ensure that conclusions drawn from data-driven approaches are robust and sound. Furthermore, Member States should welcome and support the role of the scientific community in contributing to policy, and in cultivating awareness of the strengths and weaknesses of AI technologies.

POLICY AREA 9: ECONOMY AND LABOUR

114. Member States should assess and address the impact of AI systems on labour markets and its implications for education requirements, in all countries and with special emphasis on countries

where the economy is labour-intensive. This can include the introduction of a wider range of “core” and interdisciplinary skills at all education levels to provide current workers and new generations a fair chance of finding jobs in a rapidly changing market and to ensure their awareness of the ethical aspects of AI systems. Skills such as “learning how to learn”, communication, critical thinking, teamwork, empathy, and the ability to transfer one’s knowledge across domains, should be taught alongside specialist, technical skills, as well as low-skilled tasks such as labelling datasets. Being transparent about what skills are in demand and updating curricula around these are key.

115. Member States should support collaboration agreements among governments, academic institutions, industry, workers’ organizations and civil society to bridge the gap of skillset requirements to align training programmes and strategies with the implications of the future of work and the needs of industry. Project-based teaching and learning approaches for AI should be promoted, allowing for partnerships between private sector companies, universities and research centres.

116. Member States should work with private sector companies, civil society organizations and other stakeholders, including workers and unions to ensure a fair transition for at-risk employees. This includes putting in place upskilling and reskilling programmes, finding effective mechanisms of retaining employees during those transition periods, and exploring “safety net” programmes for those who cannot be retrained. Member States should develop and implement programmes to research and address the challenges identified that could include upskilling and reskilling, enhanced social protection, proactive industry policies and interventions, tax benefits, new taxations forms, among others. Tax regimes and other relevant regulations should be carefully examined and changed if needed to counteract the consequences of unemployment caused by AI-based automation.

117. Member States should encourage and support researchers to analyse the impact of AI systems on the local labour environment in order to anticipate future trends and challenges. These studies should investigate the impact of AI systems on economic, social and geographic sectors, as well as on human-robot interactions and human-human relationships, in order to advise on reskilling and redeployment best practices.

118. Member States should devise mechanisms to prevent the monopolization of AI systems throughout their life cycle and the resulting inequalities, whether these are data, research, technology, market or other monopolies. Member States should assess relevant markets, and regulate and intervene if such monopolies exist, taking into account that, due to a lack of infrastructure, human capacity and regulations, LMICs, in particular LDCs, LLDCs and SIDS are more exposed and vulnerable to exploitation by large technology companies.

POLICY AREA 10: HEALTH AND SOCIAL WELL-BEING

119. Member States should endeavour to employ effective AI systems for improving human health and protecting the right to life, while building and maintaining international solidarity to tackle global health risks and uncertainties, and ensure that their deployment of AI systems in health care be consistent with international law and international human rights law, standards and principles. Member States should ensure that actors involved in health care AI systems take into consideration the importance of a patient’s relationships with their family and with health care staff.

120. Member States should regulate the development and deployment of AI systems related to health in general and mental health in particular to ensure that they are safe, effective, efficient, scientifically and medically sound. Moreover, in the related area of digital health interventions, Member States are strongly encouraged to actively involve patients and their representatives in all relevant steps of the development of the system.

121. Member States should pay particular attention in regulating prediction, detection and treatment solutions for health care in AI applications by:

(a) ensuring oversight to minimize bias;

(b) ensuring that the professional, the patient, caregiver or service user is included as a “domain expert” in the team when developing the algorithms;

(c) paying due attention to privacy because of the potential need of being constantly monitored;

(d) ensuring that those whose data is being analysed are aware of and provide informed consent to the tracking and analysis of their data; and

(e) ensuring the human care and final decision of diagnosis and treatment are taken by humans while acknowledging that AI systems can also assist in their work.

122. Member States should establish research on the effects and regulation of potential harms to mental health related to AI systems, such as higher degrees of depression, anxiety, social isolation, developing addiction, trafficking and radicalization, misinformation, among others.

123. Member States should develop guidelines for human-robot interactions and their impact on human-human relationships, based on research and directed at the future development of robots, with special attention to the mental and physical health of human beings, especially regarding robots in health care and the care for older persons and persons with disabilities, and regarding educational robots, toy robots, chatbots, and companion robots for children and adults. Furthermore, assistance of AI technologies should be applied to increase the safety and ergonomic use of robots, including in a human-robot working environment.

124. Member States should ensure that human-robot interactions comply with the same values and principles that apply to any other AI systems, including human rights, the promotion of diversity in relationships, and the protection of vulnerable groups.

125. Member States should protect the right of users to easily identify whether they are interacting with a living being, or with an AI system imitating human or animal characteristics.

126. Member States should implement policies to raise awareness about the anthropomorphization of AI technologies, including in the language used to mention them, and assess the manifestations, ethical implications and possible limitations of such anthropomorphization in particular in the context of robot-human interaction and especially when children are involved.

127. Member States should encourage and promote collaborative research into the effects of long-term interaction of people with AI systems, paying particular attention to the psychological and cognitive impact that these systems can have on children and young people. This should be done using multiple norms, principles, protocols, disciplinary approaches, and assessment of the modification of behaviours and habits, as well as careful evaluation of the downstream cultural and societal impacts.

128. Member States, as well as all stakeholders, should put in place mechanisms to meaningfully engage children and young people in conversations, debates, and decision-making with regards to the impact of AI systems on their lives and futures.

129. Member States should promote the accountable use of AI systems to counter hate speech in the online domain and disinformation and also to ensure that AI systems are not used to produce and spread such content, particularly in times of elections.

130. Member States should create enabling environments for media to have the rights and resources to effectively report on the benefits and harms of AI systems, and also to make use of AI systems in their reporting.

V. MONITORING AND EVALUATION

131. Member States should, according to their specific conditions, governing structures and constitutional provisions, credibly and transparently monitor and evaluate policies, programmes and mechanisms related to ethics of AI using a combination of quantitative and qualitative approaches. In support to Member States, UNESCO can contribute by:

(a) developing a globally accepted methodology for Ethical Impact Assessment (EIA) of AI technologies, including guidance for its implementation in all stages of the AI system life cycle, based on rigorous scientific research;

(b) developing a readiness methodology to assist Member States in identifying their status at specific moments of their readiness trajectory along a continuum of dimensions;

(c) developing a globally accepted methodology to evaluate *ex ante* and *ex post* the effectiveness and efficiency of the policies for AI ethics and incentives against defined objectives;

(d) strengthening the research- and evidence-based analysis of and reporting on policies regarding AI ethics, including the publication of a comparative index; and

(e) collecting and disseminating progress, innovations, research reports, scientific publications, data and statistics regarding policies for AI ethics, to support sharing best practices and mutual learning, and to advance the implementation of this Recommendation.

132. Processes for monitoring and evaluation should ensure broad participation of relevant stakeholders, including, but not limited to, people of different age groups, girls and women, persons with disabilities, disadvantaged, marginalized and vulnerable populations, indigenous communities, as well as people from diverse socio-economic backgrounds. Social, cultural, and gender diversity must be ensured, with a view to improving learning processes and strengthening the connections between findings, decision-making, transparency and accountability for results.

133. In the interests of promoting best policies and practices related to ethics of AI, appropriate tools and indicators should be developed for assessing the effectiveness and efficiency thereof against agreed standards, priorities and targets, including specific targets for persons belonging to disadvantaged, marginalized and vulnerable groups, as well as the impact of AI systems at individual and societal levels. The monitoring and assessment of the impact of AI systems and related AI ethics policies and practices should be carried out continuously in a systematic way. This should be based on internationally agreed frameworks and involve evaluations of private and public institutions, providers and programmes, including self-evaluations, as well as tracer studies and the development of sets of indicators. Data collection and processing should be conducted in accordance with national legislation on data protection and data privacy.

134. The possible mechanisms for monitoring and evaluation may include an AI ethics observatory, or contributions to existing initiatives by addressing adherence to ethical principles across UNESCO's areas of competence, an experience-sharing mechanism for Member States to provide feedback on each other's initiatives, AI regulatory sandboxes, and an assessment guide for all AI actors to evaluate their adherence to policy recommendations mentioned in this document.

VI. UTILIZATION AND EXPLOITATION OF THE PRESENT RECOMMENDATION

135. Member States and all other stakeholders as identified in this Recommendation must respect, promote and protect the ethical principles and standards regarding AI that are identified in this Recommendation, and should take all feasible steps to give effect to its policy recommendations.

136. Member States should strive to extend and complement their own action in respect of this Recommendation, by cooperating with all national and international governmental and non-governmental organizations, as well as transnational corporations and scientific organizations, whose activities fall within the scope and objectives of this Recommendation. The development of a globally accepted Ethical Impact Assessment methodology and the establishment of national commissions for the ethics of technology can be important instruments for this.

VII. PROMOTION OF THE PRESENT RECOMMENDATION

137. UNESCO has the vocation to be the principal United Nations agency to promote and disseminate this Recommendation, and accordingly shall work in collaboration with other United Nations entities, including but not limited to the United Nations Secretary-General's High-level Panel on Digital Cooperation, the World Commission on the Ethics of Scientific Knowledge and Technology (COMEST), the International Bioethics Committee (IBC), the Intergovernmental Bioethics Committee (IGBC), the International Telecommunication Union (ITU), the International Labour Organization (ILO), the World Intellectual Property Organization (WIPO), the United Nations Children's Fund (UNICEF), UN Women, the United Nations Industrial Development Organization (UNIDO), the World Trade Organization (WTO), and other relevant United Nations entities concerned with the ethics of AI.

138. UNESCO shall also work in collaboration with other international and regional organizations, including but not limited to the African Union (AU), the Alianza del Pacífico, the Association of African Universities (AAU), the Association of Southeast Asian Nations (ASEAN), the Caribbean Community (CARICOM), the Caribbean Telecommunications Union, the Caribbean Public Services Association, the Common Market for Eastern and Southern Africa (COMESA), the Community of Latin American and Caribbean States (CELAC), the Council of Europe (CoE), the Economic Community of West African States (ECOWAS), the Eurasian Economic Union (EAEU), the European Union (EU), the International Association of Universities (IAU), the Organisation for Economic Co-operation and Development (OECD), the Organization for Security and Co-operation in Europe (OSCE), the South Asian Association for Regional Cooperation (SAARC), the Southern African Development Community (SADC), the Southern Common Market (MERCOSUR), as well as the Institute of Electrical and Electronic Engineers (IEEE), the International Organization for Standardization (ISO), and international financing institutions such as the World Bank, the InterAmerican Development Bank, and the African Development Bank.

139. Even though, within UNESCO, the mandate to promote and protect falls within the authority of governments and intergovernmental bodies, civil society will be an important actor to advocate for the public sector's interests and therefore UNESCO needs to ensure and promote its legitimacy.

VIII. FINAL PROVISIONS

140. This Recommendation needs to be understood as a whole, and the foundational values and principles are to be understood as complementary and interrelated.

141. Nothing in this Recommendation may be interpreted as approval for any State, other social actor, group, or person to engage in any activity or perform any act contrary to human rights, fundamental freedoms, human dignity and concern for the environment and ecosystems.

Poder Judiciário
Conselho Nacional de Justiça

RESOLUÇÃO Nº 332, DE 21 DE AGOSTO DE 2020.

Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências.

O **PRESIDENTE DO CONSELHO NACIONAL DE JUSTIÇA**, no uso de suas atribuições legais e regimentais;

CONSIDERANDO que a Inteligência Artificial, ao ser aplicada no Poder Judiciário, pode contribuir com a agilidade e coerência do processo de tomada de decisão;

CONSIDERANDO que, no desenvolvimento e na implantação da Inteligência Artificial, os tribunais deverão observar sua compatibilidade com os Direitos Fundamentais;

CONSIDERANDO que a Inteligência Artificial aplicada nos processos de tomada de decisão deve atender a critérios éticos de transparência, previsibilidade, possibilidade de auditoria e garantia de imparcialidade e justiça substancial;

CONSIDERANDO que as decisões judiciais apoiadas pela Inteligência Artificial devem preservar a igualdade, a não discriminação, a pluralidade, a solidariedade e o julgamento justo, com a viabilização de meios destinados a eliminar ou minimizar a opressão, a marginalização do ser humano e os erros de julgamento decorrentes de preconceitos;

Poder Judiciário
Conselho Nacional de Justiça

CONSIDERANDO que os dados utilizados no processo de aprendizado de máquina deverão ser provenientes de fontes seguras, preferencialmente governamentais, passíveis de serem rastreados e auditados;

CONSIDERANDO que, no seu processo de tratamento, os dados utilizados devem ser eficazmente protegidos contra riscos de destruição, modificação, extravio, acessos e transmissões não autorizadas;

CONSIDERANDO que o uso da Inteligência Artificial deve respeitar a privacidade dos usuários, cabendo-lhes ciência e controle sobre o uso de dados pessoais;

CONSIDERANDO que os dados coletados pela Inteligência Artificial devem ser utilizados de forma responsável para proteção do usuário;

CONSIDERANDO que a utilização da Inteligência Artificial deve se desenvolver com vistas à promoção da igualdade, da liberdade e da justiça, bem como para garantir e fomentar a dignidade humana;

CONSIDERANDO o contido na Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seus ambientes;

CONSIDERANDO a ausência, no Brasil, de normas específicas quanto à governança e aos parâmetros éticos para o desenvolvimento e uso da Inteligência Artificial;

CONSIDERANDO as inúmeras iniciativas envolvendo Inteligência Artificial no âmbito do Poder Judiciário e a necessidade de observância de parâmetros para sua governança e desenvolvimento e uso éticos;

Poder Judiciário
Conselho Nacional de Justiça

CONSIDERANDO a competência do Conselho Nacional de Justiça para zelar pelo cumprimento dos princípios da administração pública no âmbito do Poder Judiciário, à exceção do Supremo Tribunal Federal, conforme art. 103- B, § 4º, II, da Constituição da República Federativa do Brasil;

CONSIDERANDO a decisão proferida pelo Plenário do Conselho Nacional de Justiça no julgamento do Procedimento de Ato Normativo nº 0005432-29.2020.2.00.0000, na 71ª Sessão Virtual, realizada em 14 de agosto de 2020;

RESOLVE:

CAPÍTULO I
DAS DISPOSIÇÕES GERAIS

Art. 1º O conhecimento associado à Inteligência Artificial e a sua implementação estarão à disposição da Justiça, no sentido de promover e aprofundar maior compreensão entre a lei e o agir humano, entre a liberdade e as instituições judiciais.

Art. 2º A Inteligência Artificial, no âmbito do Poder Judiciário, visa promover o bem-estar dos jurisdicionados e a prestação equitativa da jurisdição, bem como descobrir métodos e práticas que possibilitem a consecução desses objetivos.

Art. 3º Para o disposto nesta Resolução, considera-se:

I – Algoritmo: sequência finita de instruções executadas por um programa de computador, com o objetivo de processar informações para um fim específico;

II – Modelo de Inteligência Artificial: conjunto de dados e algoritmos computacionais, concebidos a partir de modelos matemáticos, cujo objetivo é oferecer resultados inteligentes, associados ou comparáveis a determinados aspectos do pensamento, do saber ou da atividade humana;

III – Sinapses: solução computacional, mantida pelo Conselho Nacional de Justiça, com o objetivo de armazenar, testar, treinar, distribuir e auditar modelos de Inteligência Artificial;

IV – Usuário: pessoa que utiliza o sistema inteligente e que tem direito ao seu controle, conforme sua posição endógena ou exógena ao Poder Judiciário, pode ser um usuário interno ou um usuário externo;

V – Usuário interno: membro, servidor ou colaborador do Poder Judiciário que desenvolva ou utilize o sistema inteligente;

VI – Usuário externo: pessoa que, mesmo sem ser membro, servidor ou colaborador do Poder Judiciário, utiliza ou mantém qualquer espécie de contato com o sistema inteligente, notadamente jurisdicionados, advogados, defensores públicos, procuradores, membros do Ministério Público, peritos, assistentes técnicos, entre outros.

CAPÍTULO II
DO RESPEITO AOS DIREITOS FUNDAMENTAIS

Art. 4º No desenvolvimento, na implantação e no uso da Inteligência Artificial, os tribunais observarão sua compatibilidade com os Direitos Fundamentais, especialmente aqueles previstos na Constituição ou em tratados de que a República Federativa do Brasil seja parte.

Art. 5º A utilização de modelos de Inteligência Artificial deve buscar garantir a segurança jurídica e colaborar para que o Poder Judiciário respeite a igualdade de tratamento aos casos absolutamente iguais.

Art. 6º Quando o desenvolvimento e treinamento de modelos de Inteligência exigir a utilização de dados, as amostras devem ser representativas e observar as cautelas necessárias quanto aos dados pessoais sensíveis e ao segredo de justiça.

Parágrafo único. Para fins desta Resolução, são dados pessoais sensíveis aqueles assim considerados pela Lei nº 13.709/2018, e seus atos regulamentares.

CAPÍTULO III
DA NÃO DISCRIMINAÇÃO

Art. 7º As decisões judiciais apoiadas em ferramentas de Inteligência Artificial devem preservar a igualdade, a não discriminação, a pluralidade e a solidariedade, auxiliando no julgamento justo, com criação de condições que visem eliminar ou minimizar a opressão, a marginalização do ser humano e os erros de julgamento decorrentes de preconceitos.

§ 1º Antes de ser colocado em produção, o modelo de Inteligência Artificial deverá ser homologado de forma a identificar se preconceitos ou generalizações influenciaram seu desenvolvimento, acarretando tendências discriminatórias no seu funcionamento.

§ 2º Verificado viés discriminatório de qualquer natureza ou incompatibilidade do modelo de Inteligência Artificial com os princípios previstos nesta Resolução, deverão ser adotadas medidas corretivas.

§ 3º A impossibilidade de eliminação do viés discriminatório do modelo de Inteligência Artificial implicará na descontinuidade de sua utilização, com o consequente registro de seu projeto e as razões que levaram a tal decisão.

CAPÍTULO IV
DA PUBLICIDADE E TRANSPARÊNCIA

Art. 8º Para os efeitos da presente Resolução, transparência consiste em:

I – divulgação responsável, considerando a sensibilidade própria dos dados judiciais;

II – indicação dos objetivos e resultados pretendidos pelo uso do modelo de Inteligência Artificial;

III – documentação dos riscos identificados e indicação dos instrumentos de segurança da informação e controle para seu enfrentamento;

IV – possibilidade de identificação do motivo em caso de dano causado pela ferramenta de Inteligência Artificial;

V – apresentação dos mecanismos de auditoria e certificação de boas práticas;

VI – fornecimento de explicação satisfatória e passível de auditoria por autoridade humana quanto a qualquer proposta de decisão apresentada pelo modelo de Inteligência Artificial, especialmente quando essa for de natureza judicial.

CAPÍTULO V
DA GOVERNANÇA E DA QUALIDADE

Art. 9º Qualquer modelo de Inteligência Artificial que venha a ser adotado pelos órgãos do Poder Judiciário deverá observar as regras de governança de dados aplicáveis aos seus próprios sistemas computacionais, as Resoluções e as Recomendações do Conselho Nacional de Justiça, a Lei nº 13.709/2018, e o segredo de justiça.

Art. 10. Os órgãos do Poder Judiciário envolvidos em projeto de Inteligência Artificial deverão:

I – informar ao Conselho Nacional de Justiça a pesquisa, o desenvolvimento, a implantação ou o uso da Inteligência Artificial, bem como os respectivos objetivos e os resultados que se pretende alcançar;

II – promover esforços para atuação em modelo comunitário, com vedação a desenvolvimento paralelo quando a iniciativa possuir objetivos e resultados alcançados idênticos a modelo de Inteligência Artificial já existente ou com projeto em andamento;

III – depositar o modelo de Inteligência Artificial no Sinapses.

Art. 11. O Conselho Nacional de Justiça publicará, em área própria de seu sítio na rede mundial de computadores, a relação dos modelos de Inteligência Artificial desenvolvidos ou utilizados pelos órgãos do Poder Judiciário.

Art. 12. Os modelos de Inteligência Artificial desenvolvidos pelos órgãos do Poder Judiciário deverão possuir interface de programação de aplicativos (API) que permitam sua utilização por outros sistemas.

Parágrafo único. O Conselho Nacional de Justiça estabelecerá o padrão de interface de programação de aplicativos (API) mencionado no *caput* deste artigo.

CAPÍTULO VI
DA SEGURANÇA

Art. 13. Os dados utilizados no processo de treinamento de modelos de Inteligência Artificial deverão ser provenientes de fontes seguras, preferencialmente governamentais.

Art. 14. O sistema deverá impedir que os dados recebidos sejam alterados antes de sua utilização nos treinamentos dos modelos, bem como seja mantida sua cópia (dataset) para cada versão de modelo desenvolvida.

Art. 15. Os dados utilizados no processo devem ser eficazmente protegidos contra os riscos de destruição, modificação, extravio ou acessos e transmissões não autorizados.

Art. 16. O armazenamento e a execução dos modelos de Inteligência Artificial deverão ocorrer em ambientes aderentes a padrões consolidados de segurança da informação.

CAPÍTULO VII
DO CONTROLE DO USUÁRIO

Art. 17. O sistema inteligente deverá assegurar a autonomia dos usuários internos, com uso de modelos que:

I – proporcione incremento, e não restrição;

II – possibilite a revisão da proposta de decisão e dos dados utilizados para sua elaboração, sem que haja qualquer espécie de vinculação à solução apresentada pela Inteligência Artificial.

Art. 18. Os usuários externos devem ser informados, em linguagem clara e precisa, quanto à utilização de sistema inteligente nos serviços que lhes forem prestados.

Parágrafo único. A informação prevista no *caput* deve destacar o caráter não vinculante da proposta de solução apresentada pela Inteligência Artificial, a qual sempre é submetida à análise da autoridade competente.

Art. 19. Os sistemas computacionais que utilizem modelos de Inteligência Artificial como ferramenta auxiliar para a elaboração de decisão judicial observarão, como critério preponderante para definir a técnica utilizada, a explicação dos passos que conduziram ao resultado.

Parágrafo único. Os sistemas computacionais com atuação indicada no *caput* deste artigo deverão permitir a supervisão do magistrado competente.

CAPÍTULO VIII
DA PESQUISA, DO DESENVOLVIMENTO E DA IMPLANTAÇÃO DE SERVIÇOS DE INTELIGÊNCIA ARTIFICIAL

Art. 20. A composição de equipes para pesquisa, desenvolvimento e implantação das soluções computacionais que se utilizem de Inteligência Artificial será orientada pela busca da diversidade em seu mais amplo espectro, incluindo gênero, raça, etnia, cor, orientação sexual, pessoas com deficiência, geração e demais características individuais.

§ 1º A participação representativa deverá existir em todas as etapas do processo, tais como planejamento, coleta e processamento de dados, construção, verificação, validação e implementação dos modelos, tanto nas áreas técnicas como negociais.

§ 2º A diversidade na participação prevista no *caput* deste artigo apenas será dispensada mediante decisão fundamentada, dentre outros motivos, pela ausência de profissionais no quadro de pessoal dos tribunais.

§ 3º As vagas destinadas à capacitação na área de Inteligência Artificial serão, sempre que possível, distribuídas com observância à diversidade.

§ 4º A formação das equipes mencionadas no *caput* deverá considerar seu caráter interdisciplinar, incluindo profissionais de Tecnologia da Informação e de outras áreas cujo conhecimento científico possa contribuir para pesquisa, desenvolvimento ou implantação do sistema inteligente.

Art. 21. A realização de estudos, pesquisas, ensino e treinamentos de Inteligência Artificial deve ser livre de preconceitos, sendo vedado:

I – desrespeitar a dignidade e a liberdade de pessoas ou grupos envolvidos em seus trabalhos;

II – promover atividades que envolvam qualquer espécie de risco ou prejuízo aos seres humanos e à equidade das decisões;

III – subordinar investigações a sectarismo capaz de direcionar o curso da pesquisa ou seus resultados.

Art. 22. Iniciada pesquisa, desenvolvimento ou implantação de modelos de Inteligência Artificial, os tribunais deverão comunicar imediatamente ao Conselho Nacional de Justiça e velar por sua continuidade.

§ 1º As atividades descritas no *caput* deste artigo serão encerradas quando, mediante manifestação fundamentada, for reconhecida sua desconformidade com os preceitos éticos estabelecidos nesta Resolução ou em outros atos normativos aplicáveis ao Poder Judiciário e for inviável sua readequação.

§ 2º Não se enquadram no *caput* deste artigo a utilização de modelos de Inteligência Artificial que utilizem técnicas de reconhecimento facial, os quais exigirão prévia autorização do Conselho Nacional de Justiça para implementação.

Art. 23. A utilização de modelos de Inteligência Artificial em matéria penal não deve ser estimulada, sobretudo com relação à sugestão de modelos de decisões preditivas.

§ 1º Não se aplica o disposto no *caput* quando se tratar de utilização de soluções computacionais destinadas à automação e ao oferecimento de subsídios destinados ao cálculo de penas, prescrição, verificação de reincidência, mapeamentos, classificações e triagem dos autos para fins de gerenciamento de acervo.

§ 2º Os modelos de Inteligência Artificial destinados à verificação de reincidência penal não devem indicar conclusão mais prejudicial ao réu do que aquela a que o magistrado chegaria sem sua utilização.

Art. 24. Os modelos de Inteligência Artificial utilizarão preferencialmente software de código aberto que:

I – facilite sua integração ou interoperabilidade entre os sistemas utilizados pelos órgãos do Poder Judiciário;

II – possibilite um ambiente de desenvolvimento colaborativo;

III – permita maior transparência;

IV – proporcione cooperação entre outros segmentos e áreas do setor público e a sociedade civil.

CAPÍTULO IX
DA PRESTAÇÃO DE CONTAS E DA RESPONSABILIZAÇÃO

Art. 25. Qualquer solução computacional do Poder Judiciário que utilizar modelos de Inteligência Artificial deverá assegurar total transparência na prestação de contas, com o fim de garantir o impacto positivo para os usuários finais e para a sociedade.

Parágrafo único. A prestação de contas compreenderá:

I – os nomes dos responsáveis pela execução das ações e pela prestação de contas;

II – os custos envolvidos na pesquisa, desenvolvimento, implantação, comunicação e treinamento;

III – a existência de ações de colaboração e cooperação entre os agentes do setor público ou desses com a iniciativa privada ou a sociedade civil;

IV – os resultados pretendidos e os que foram efetivamente alcançados;

V – a demonstração de efetiva publicidade quanto à natureza do serviço oferecido, técnicas utilizadas, desempenho do sistema e riscos de erros.

Art. 26. O desenvolvimento ou a utilização de sistema inteligente em desconformidade aos princípios e regras estabelecidos nesta Resolução será objeto de apuração e, sendo o caso, punição dos responsáveis.

Art. 27. Os órgãos do Poder Judiciário informarão ao Conselho Nacional de Justiça todos os registros de eventos adversos no uso da Inteligência Artificial.

CAPÍTULO X
DAS DISPOSIÇÕES FINAIS

Art. 28. Os órgãos do Poder Judiciário poderão realizar cooperação técnica
com outras instituições, públicas ou privadas, ou sociedade civil, para o desenvolvimento colaborativo de modelos de Inteligência Artificial, observadas as disposições contidas nesta Resolução, bem como a proteção dos dados que venham a ser utilizados.

Art. 29. As normas previstas nesta Resolução não excluem a aplicação de outras integrantes do ordenamento jurídico pátrio, inclusive por incorporação de tratado ou convenção internacional de que a República Federativa do Brasil seja parte.

Art. 30. As disposições desta Resolução aplicam-se inclusive aos projetos e modelos de Inteligência Artificial já em desenvolvimento ou implantados nos tribunais, respeitados os atos já aperfeiçoados.

Art. 31. Esta Resolução entra em vigor na data da sua publicação.

Ministro **DIAS TOFFOLI**