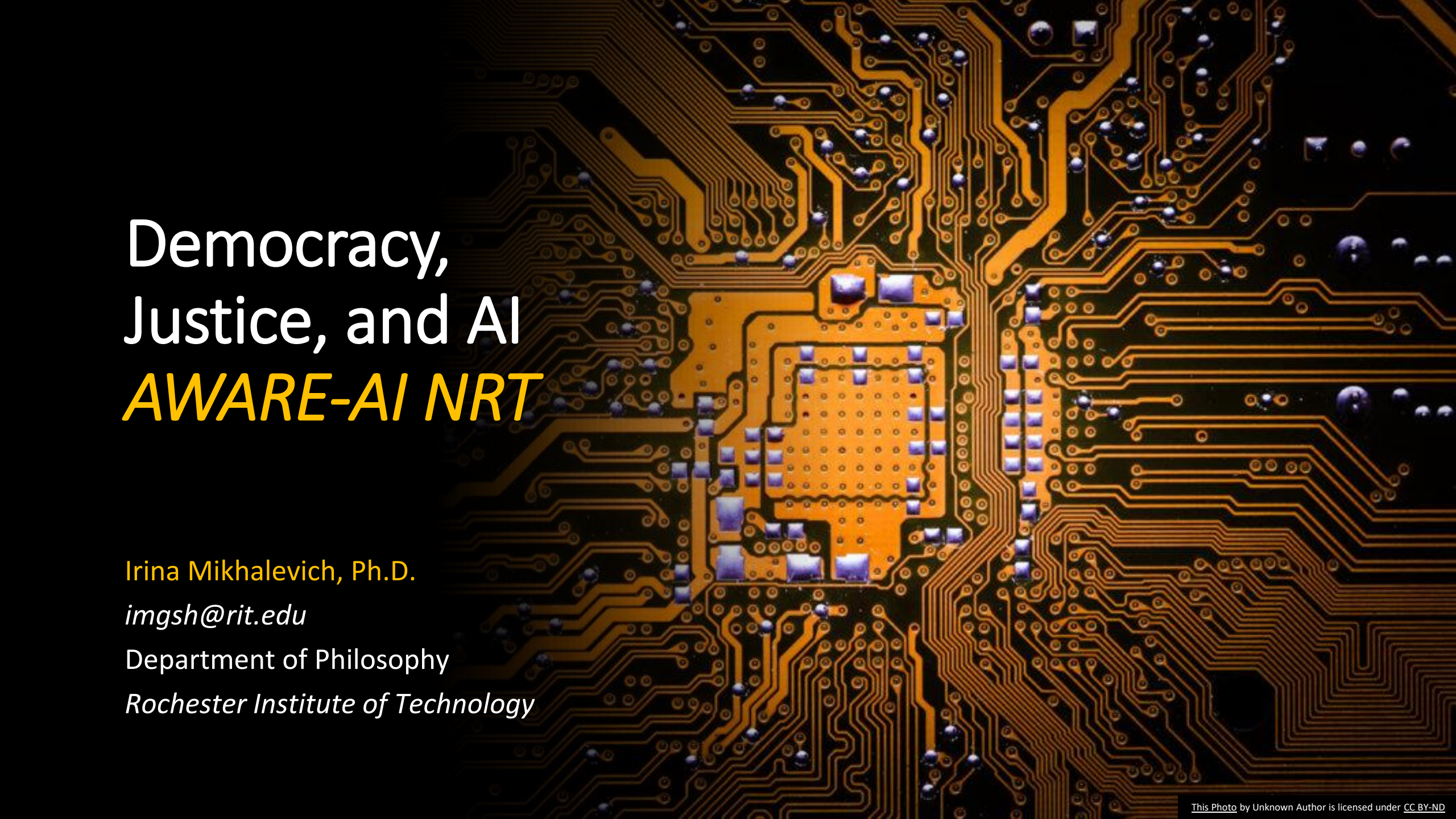
Democracy, Justice, and AI
AWARE-AI NRT

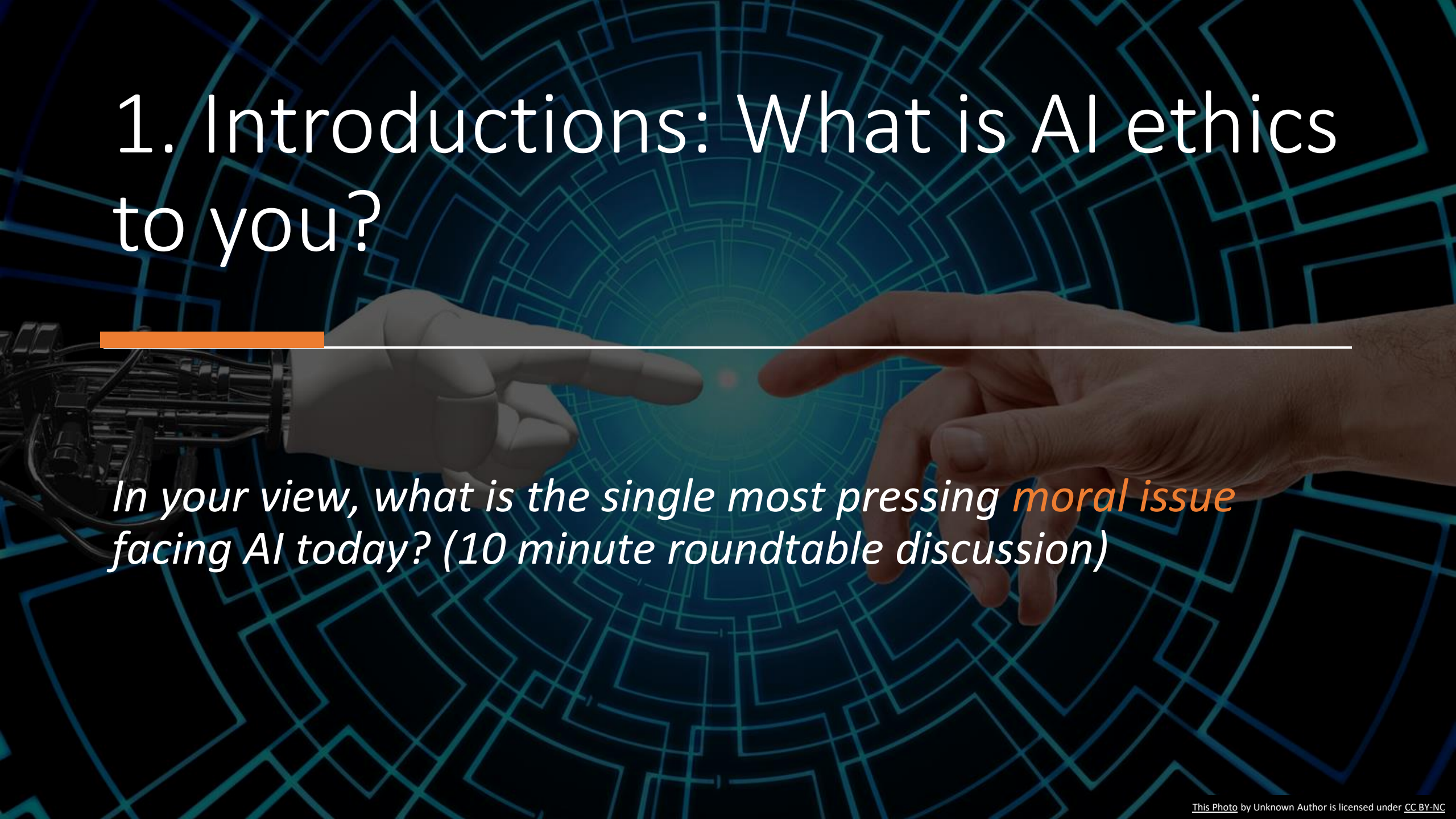
Irina Mikhalevich, Ph.D.
imgsh@rit.edu
Department of Philosophy
Rochester Institute of Technology



Outline of workshop

1. Introductions & pre-workshop writing exercise
2. Brief introduction to AI ethics
3. *Break*
4. Democratizing AI mini-lecture & discussion
5. *Break*
6. AGI & Democratic participation mini-lecture & discussion

1. Introductions: What is AI ethics to you?



In your view, what is the single most pressing moral issue facing AI today? (10 minute roundtable discussion)

2. Very Brief Introduction to AI Ethics

A broad overview and summary of one framework for ethical AI

AI Ethics as a discipline

Ethics is the study of morality.

Applied ethics is the study of morality within specific domains of life (e.g., biology, medicine, animal rights, technology, etc.)

AI Ethics is a branch of applied ethics and intersects with political philosophy, STS, cognitive science, computing, etc..

What **morality** is … and is *not*

The Minimum Conception of Morality: Morality is normative, reason-responsive, impartial, and universal (Rachells 1998).

Contrast with "thick" conceptions of morality – utilitarianism, Kantian deontology, etc.

Sources of Normativity outside of Morality:

- The law
- Cultural norms
- Religion

What do we mean by *Artificial Intelligence*?

Strong AI (~AGI) versus Weak AI (Searle 1989)

- **Strong AI & Artificial General Intelligence (AGI):** Artificial minds
  - AGI: understanding/semantics V mere syntax
  - Strong AI (Searle): AGI + consciousness
- **Weak AI:** every other form of AI and machine learning

Ethical implications of the distinction

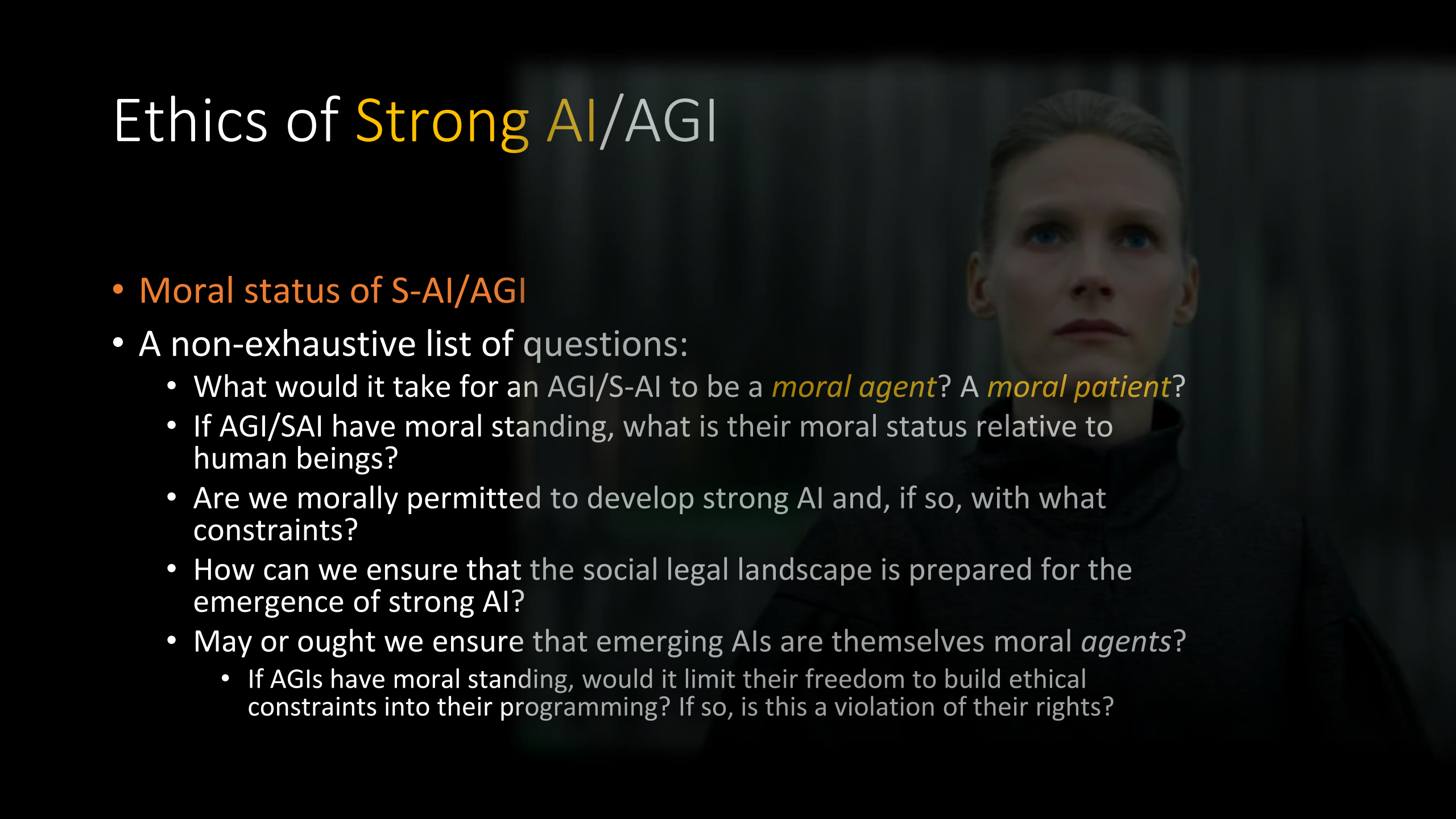
- IF strong AI/AGI → we may have moral obligations *toward the AIs themselves (including, perhaps, to involve them in democratic decision-making).*
- IF weak AI → presumably no obligations toward them. Instead, our obligations will be toward those moral patients and moral agents who would be impacted by the AI (present and future individuals or groups, the planet, etc.). Appropriate analysis would focus on threat-assessment and benefit estimates of AI vis-à-vis valuable social institutions.

Ethics of "weak" AI

Assessments of the projected **risks**, **benefits**, and **opportunity costs** associated with decisions to **invest in**, **develop**, **bring to market**, and **regulate** AI technologies must include the *moral* risks, benefits, opportunity costs.

Who are the stakeholders and with whom do the moral responsibilities lie?

Ethics of Strong AI/AGI



- Moral status of S-AI/AGI
- A non-exhaustive list of questions:
  - What would it take for an AGI/S-AI to be a *moral agent*? A *moral patient*?
  - If AGI/SAI have moral standing, what is their moral status relative to human beings?
  - Are we morally permitted to develop strong AI and, if so, with what constraints?
  - How can we ensure that the social legal landscape is prepared for the emergence of strong AI?
  - May or ought we ensure that emerging AIs are themselves moral *agents*?
    - If AGIs have moral standing, would it limit their freedom to build ethical constraints into their programming? If so, is this a violation of their rights?

A proposed framework for ethical AI

- Many proposals for ethical frameworks for the development, distribution, and regulation or AI
- We'll take a look at one: Floridi, Luciano, et al. "AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations." Minds and Machines 28.4 (2018): 689-707.
- Rationale:
    1. Synthesizes existing proposals for ethical AI,
    2. incorporates these into established bioethical paradigm,
    3. and offers concrete recommendations for inter-governmental groups (EU).

Floridi et al. (2018) propose the following ethical framing

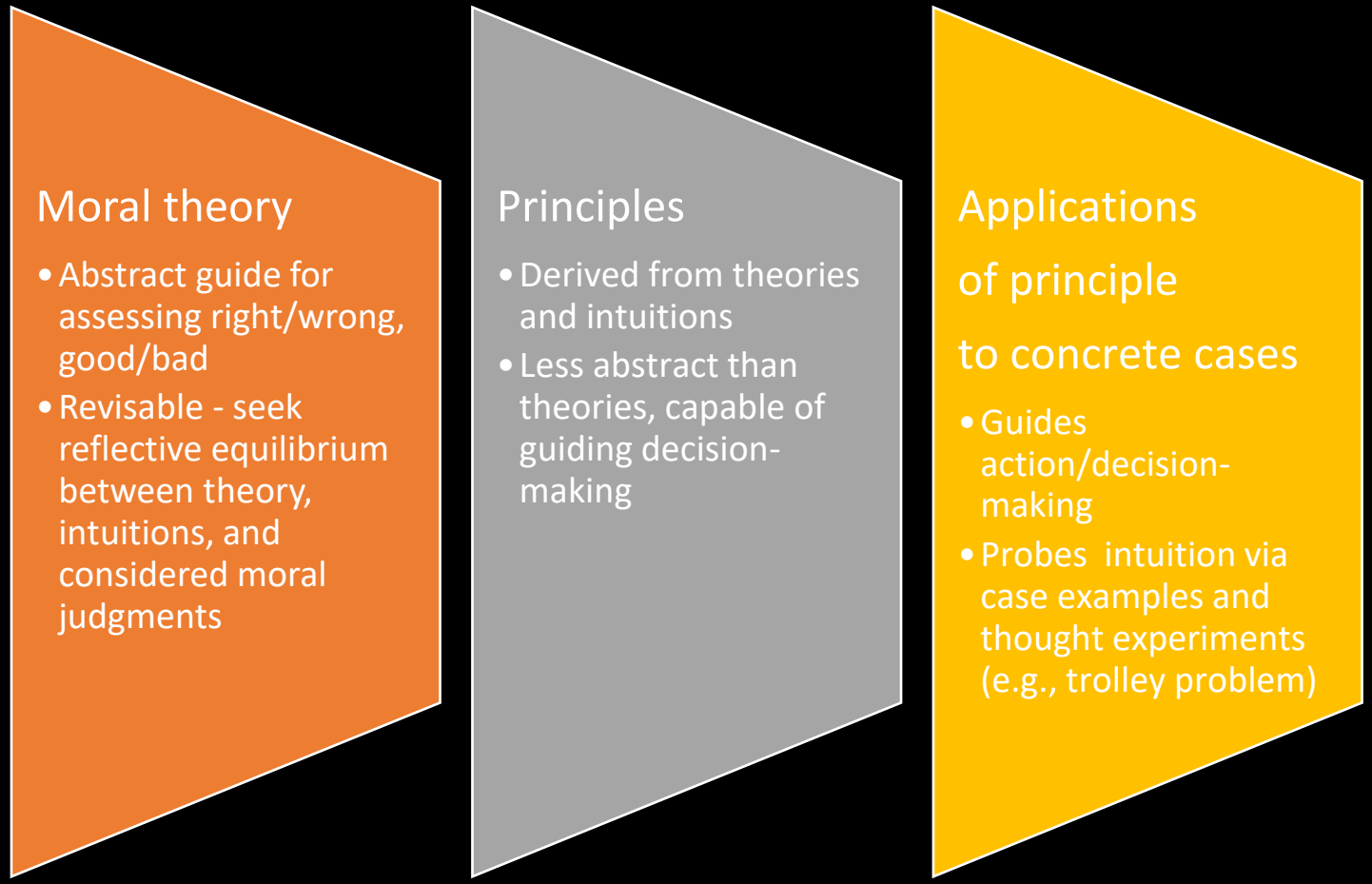
Ethical AI should be formulated with a view to preserving **human dignity** & supporting **human flourishing** (Floridi et al., 690)

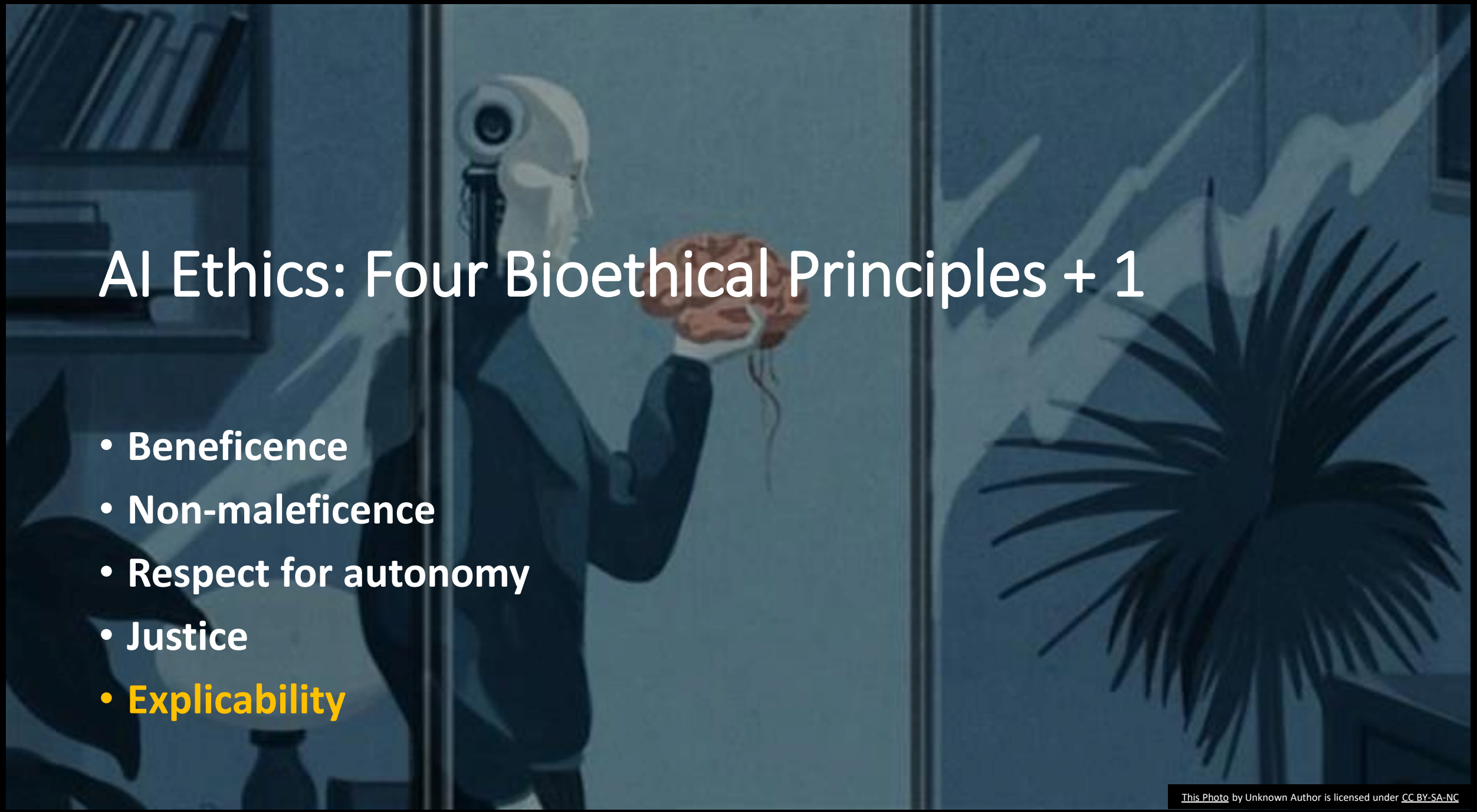
The guide toward these goals are the bioethical principles + 1 additional principle.

What are moral principles? *Where do they come from? Why should we accept them?*

Four Bioethical Principles

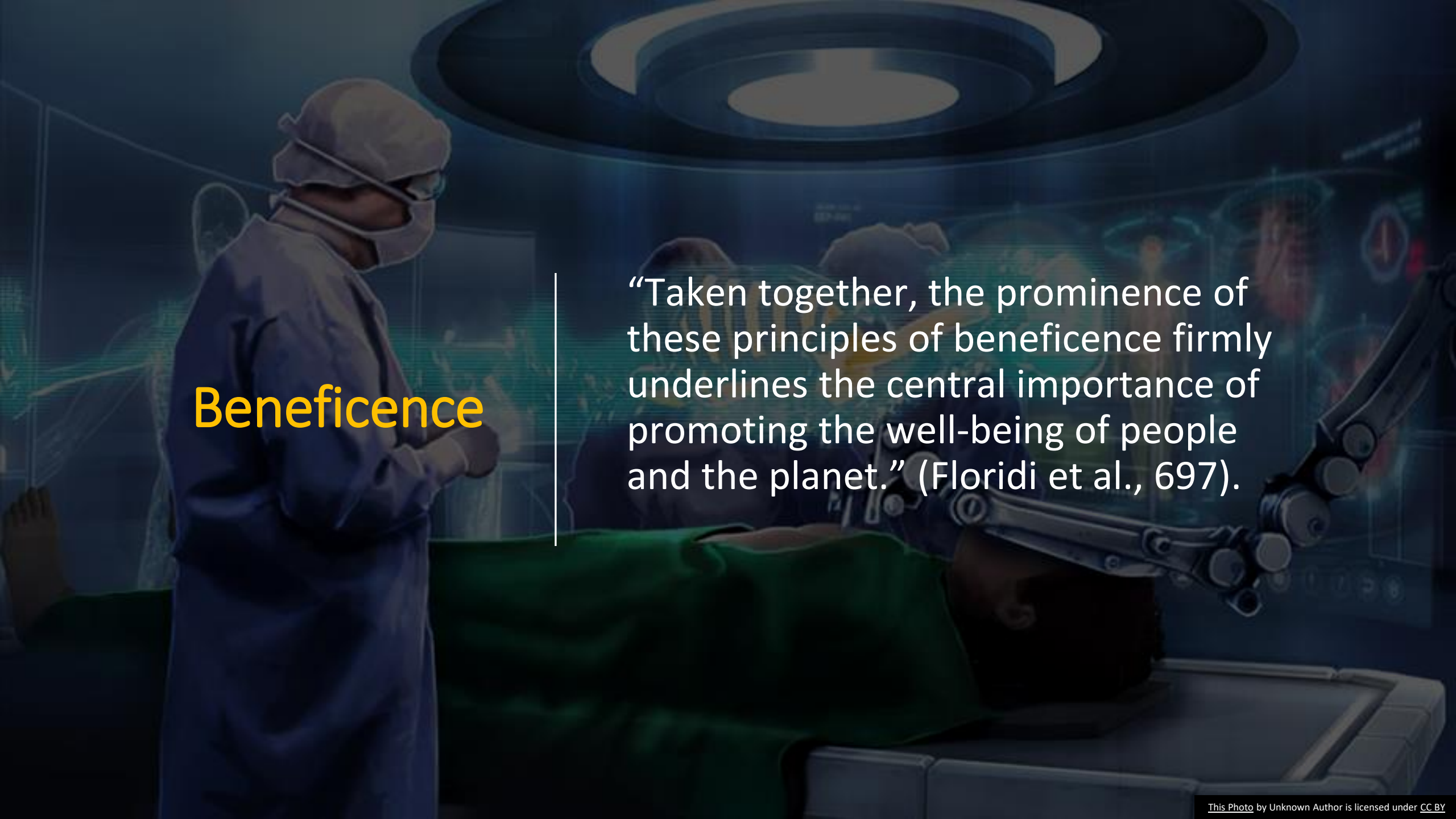
- Beneficence
- Non-maleficence
- Respect for autonomy
- Justice





AI Ethics: Four Bioethical Principles + 1

- Beneficence
- Non-maleficence
- Respect for autonomy
- Justice
- Explicability



Beneficence

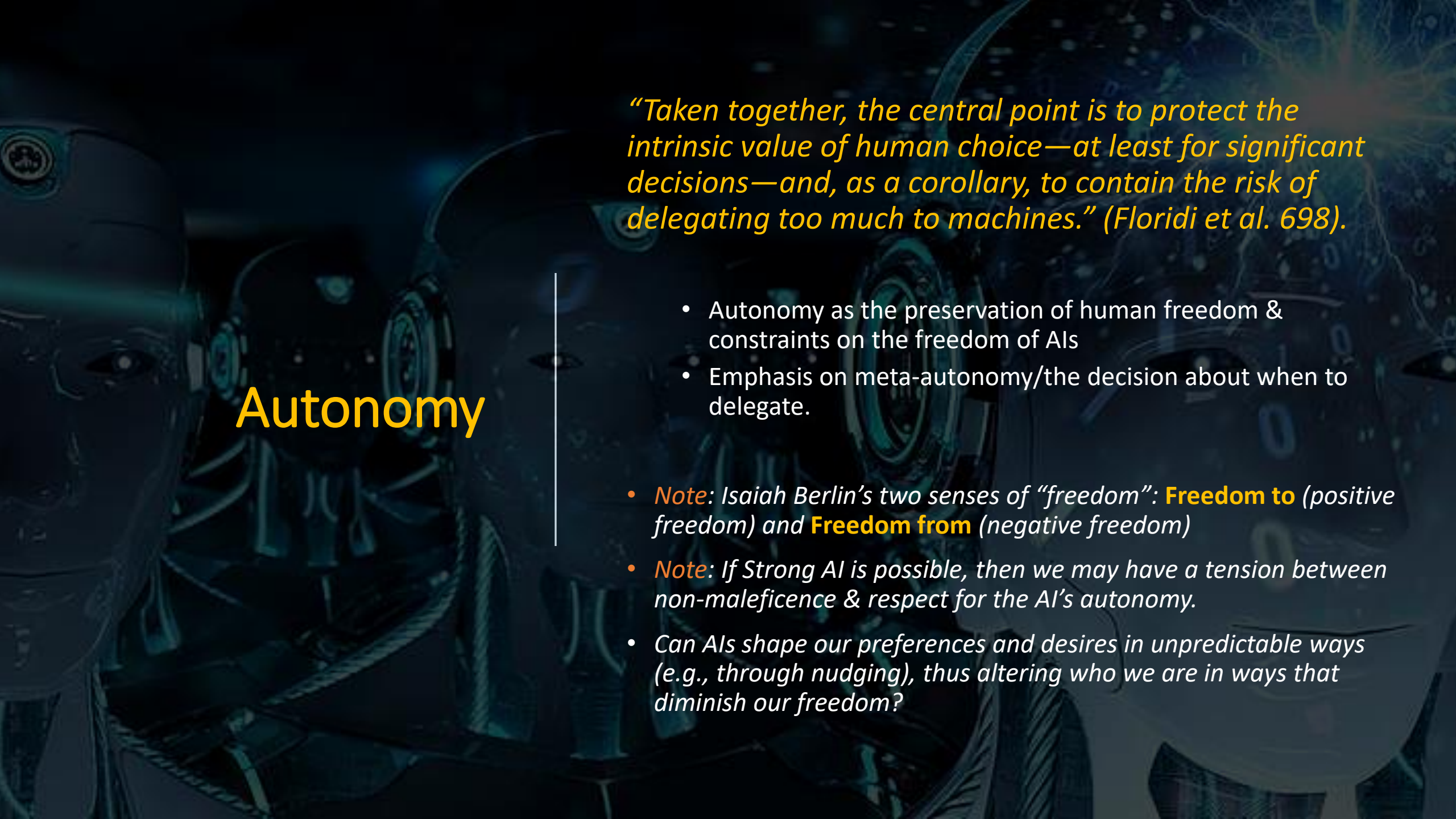
"Taken together, the prominence of these principles of beneficence firmly underlines the central importance of promoting the well-being of people and the planet." (Floridi et al., 697).

Non-maleficence

Prevent harm either "from the intent of humans or the or the unpredicted behavior of machines" (697)

Potential concerns include:

- infringements on privacy
  - "privacy is characterised as being intimately linked to individuals' access to, and control over, how personal data is used." (Floridi et al., 697)
  - Ex: facial recognition technologies & government surveillance; Evan Selinger: the right to anonymity
- Ensuring "upper limits of future to AI capabilities" (caution against recursive self-improvements)
- Spread of misinformation, disinformation
- Create or support social division, polarization, extremism, & stochastic violence

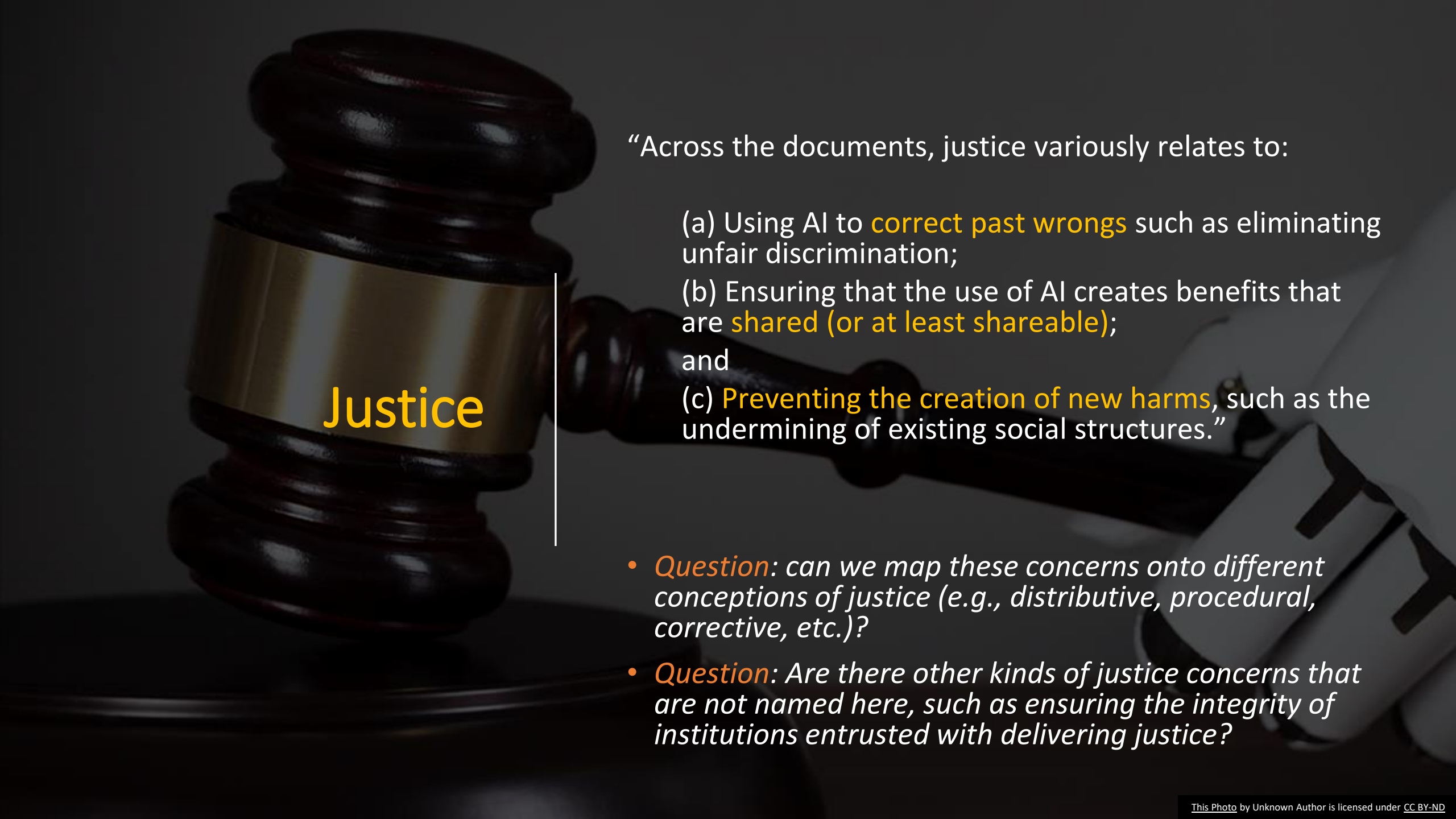


Autonomy

"Taken together, the central point is to protect the intrinsic value of human choice—at least for significant decisions—and, as a corollary, to contain the risk of delegating too much to machines." (Floridi et al. 698).

- Autonomy as the preservation of human freedom & constraints on the freedom of AIs
- Emphasis on meta-autonomy/the decision about when to delegate.

- *Note: Isaiah Berlin's two senses of "freedom": **Freedom to** (positive freedom) and **Freedom from** (negative freedom)*
- *Note: If Strong AI is possible, then we may have a tension between non-maleficence & respect for the AI's autonomy.*
- *Can AIs shape our preferences and desires in unpredictable ways (e.g., through nudging), thus altering who we are in ways that diminish our freedom?*



Justice

"Across the documents, justice variously relates to:

(a) Using AI to correct past wrongs such as eliminating unfair discrimination;

(b) Ensuring that the use of AI creates benefits that are shared (or at least shareable);

and

(c) Preventing the creation of new harms, such as the undermining of existing social structures."

- *Question: can we map these concerns onto different conceptions of justice (e.g., distributive, procedural, corrective, etc.)?*
- *Question: Are there other kinds of justice concerns that are not named here, such as ensuring the integrity of institutions entrusted with delivering justice?*

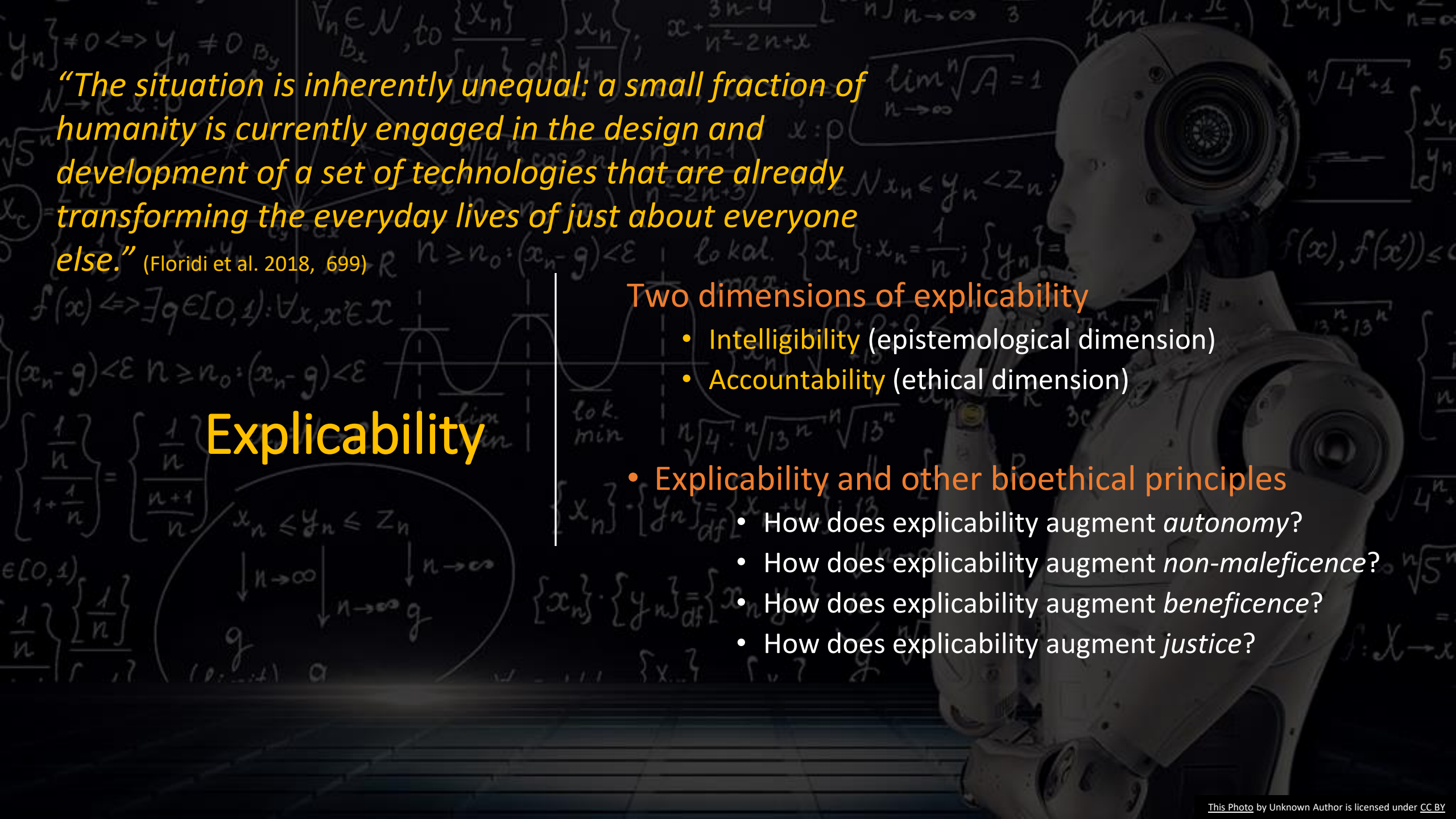
"The situation is inherently unequal: a small fraction of humanity is currently engaged in the design and development of a set of technologies that are already transforming the everyday lives of just about everyone else." (Floridi et al. 2018, 699)

Explicability

Two dimensions of explicability

- Intelligibility (epistemological dimension)
- Accountability (ethical dimension)

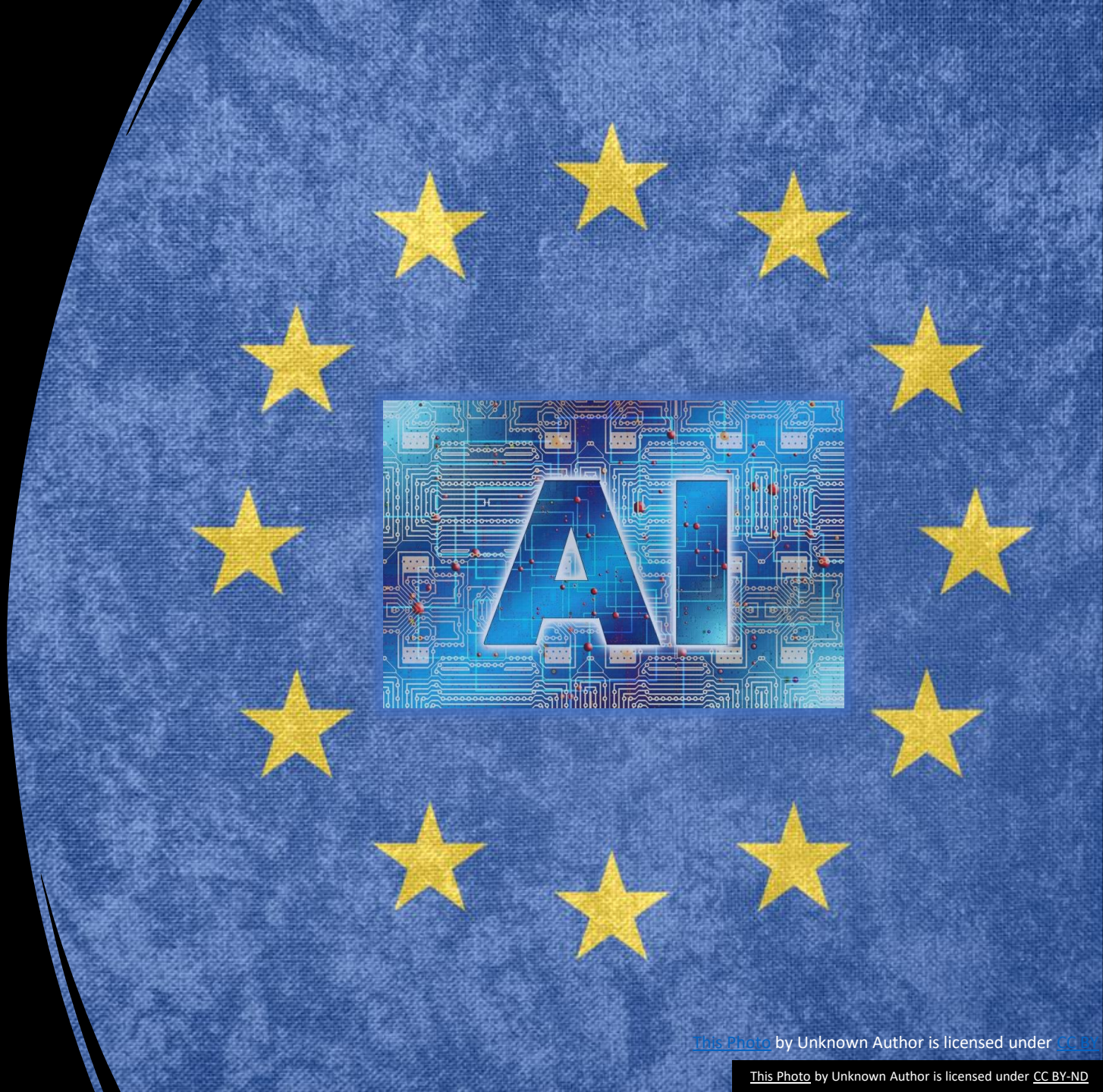
- Explicability and other bioethical principles
  - How does explicability augment *autonomy*?
  - How does explicability augment *non-maleficence*?
  - How does explicability augment *beneficence*?
  - How does explicability augment *justice*?



Summary of Recommended Actions at the level of international bodies (EU)

Four action points:

- Assessment
- Development
- Incentivization
- Support



Summing Up

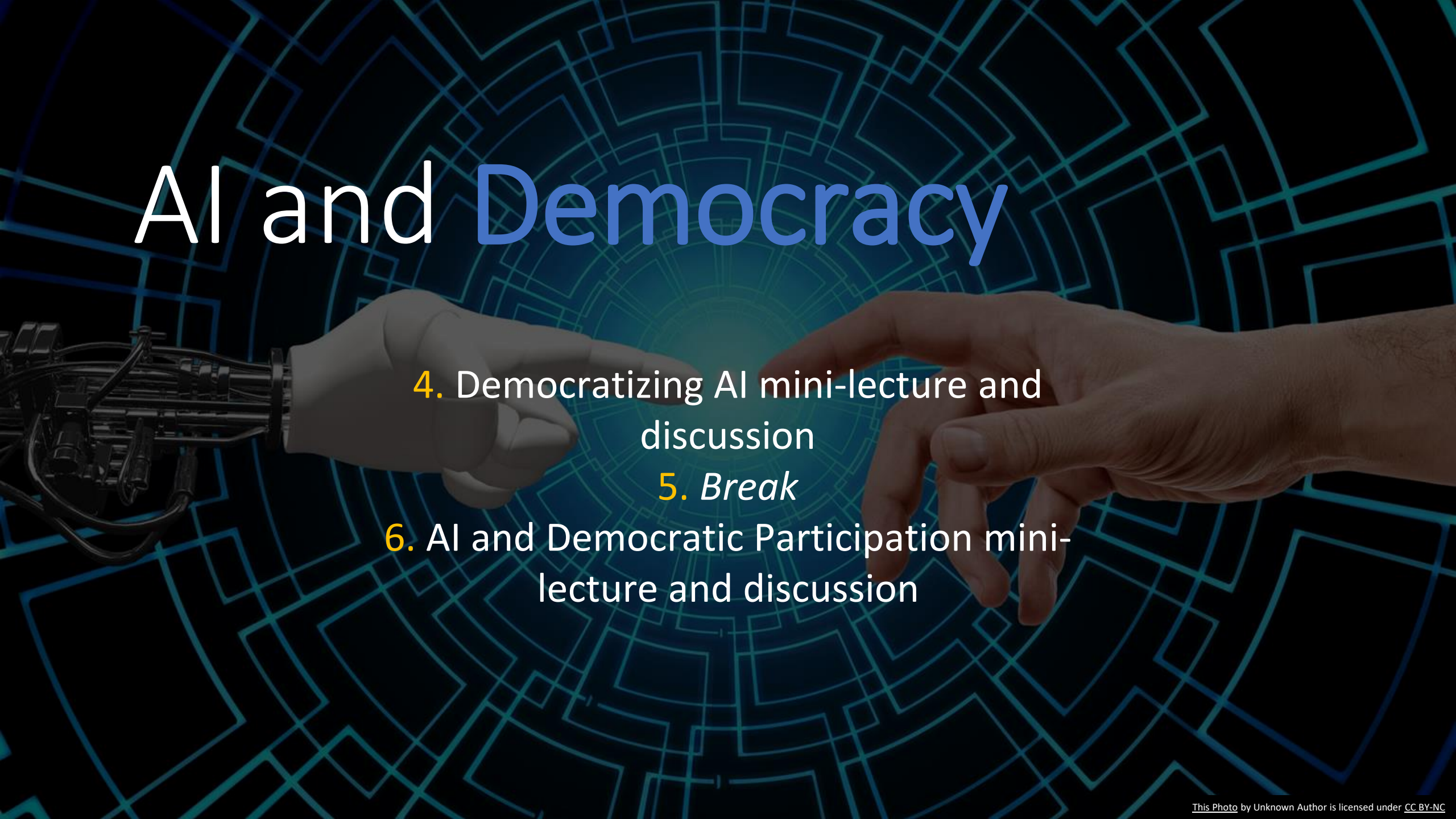
- Motivations for ethical AI: opportunities for human dignity & flourishing
- Guide should be: five ethical principles.
- Proposal: Structuring future AI assessment, development, incentivization, and support at the level of nations and multinational organizations, such as the EU with an eye to the five principles.

3-minute exercises

- *Can you reframe the concerns with which we started out this session in terms of the five bioethical principles and their relationship to human dignity and human flourishing?*
- *What, if anything, might human dignity and human flourishing have to do with democracy?*

10 minute BREAK

AI and Democracy



4. Democratizing AI mini-lecture and discussion
5. *Break*
6. AI and Democratic Participation mini-lecture and discussion

There is no question that artificial intelligence (AI) is rapidly evolving and growing more sophisticated every day. With this in mind, it is natural to wonder how AI will impact democracy.

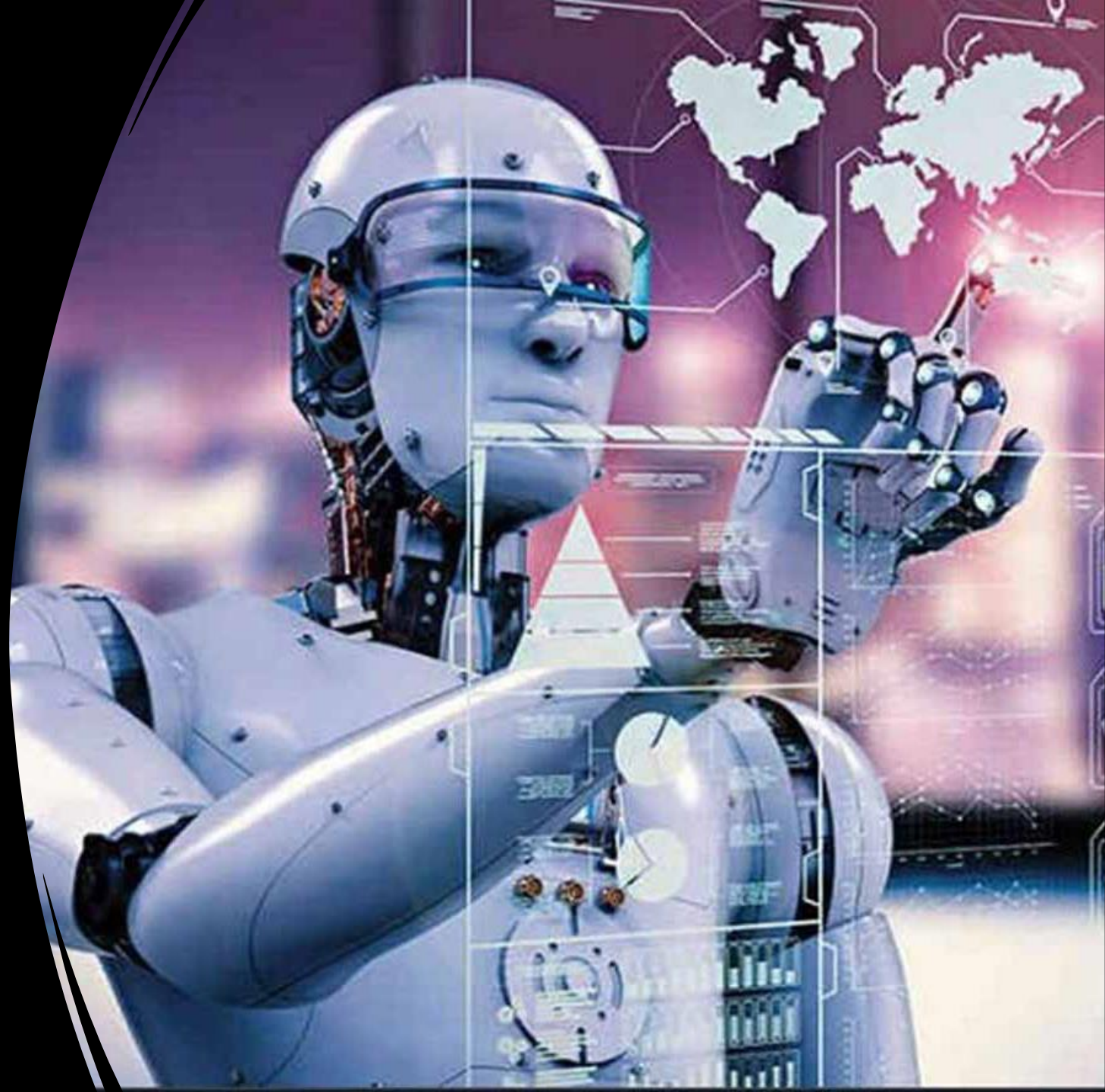
There are many potential implications of AI on democracy, both positive and negative. On the positive side, AI could help to improve the efficiency of democratic institutions and processes. For example, AI could be used to streamline the process of voter registration or to help keep track of election results. AI could also be used to help identify and track public opinion on various issues. On the negative side, AI could also be used to manipulate public opinion or to interfere with democratic processes. For example, AI could be used to spread false information or to target individuals with personalized messages designed to sway their opinions. AI could also be used to hack into election systems and tamper with the results.

Overall, it is difficult to predict exactly how AI will impact democracy. However, it is clear that AI will likely have both positive and negative implications for democracy. It is important to keep a close eye on how AI develops and to be aware of the potential risks and challenges that it poses to democracy.

- GPT-3

Strong AI/AGI & Democracy

- Do S-AIs or AGIs have a right to participate in political processes & democratic self-governance?
- Do AGIs threaten democratic institutions or processes? If so, how might their interests we weighted against human rights?
- *(More on this shortly)*



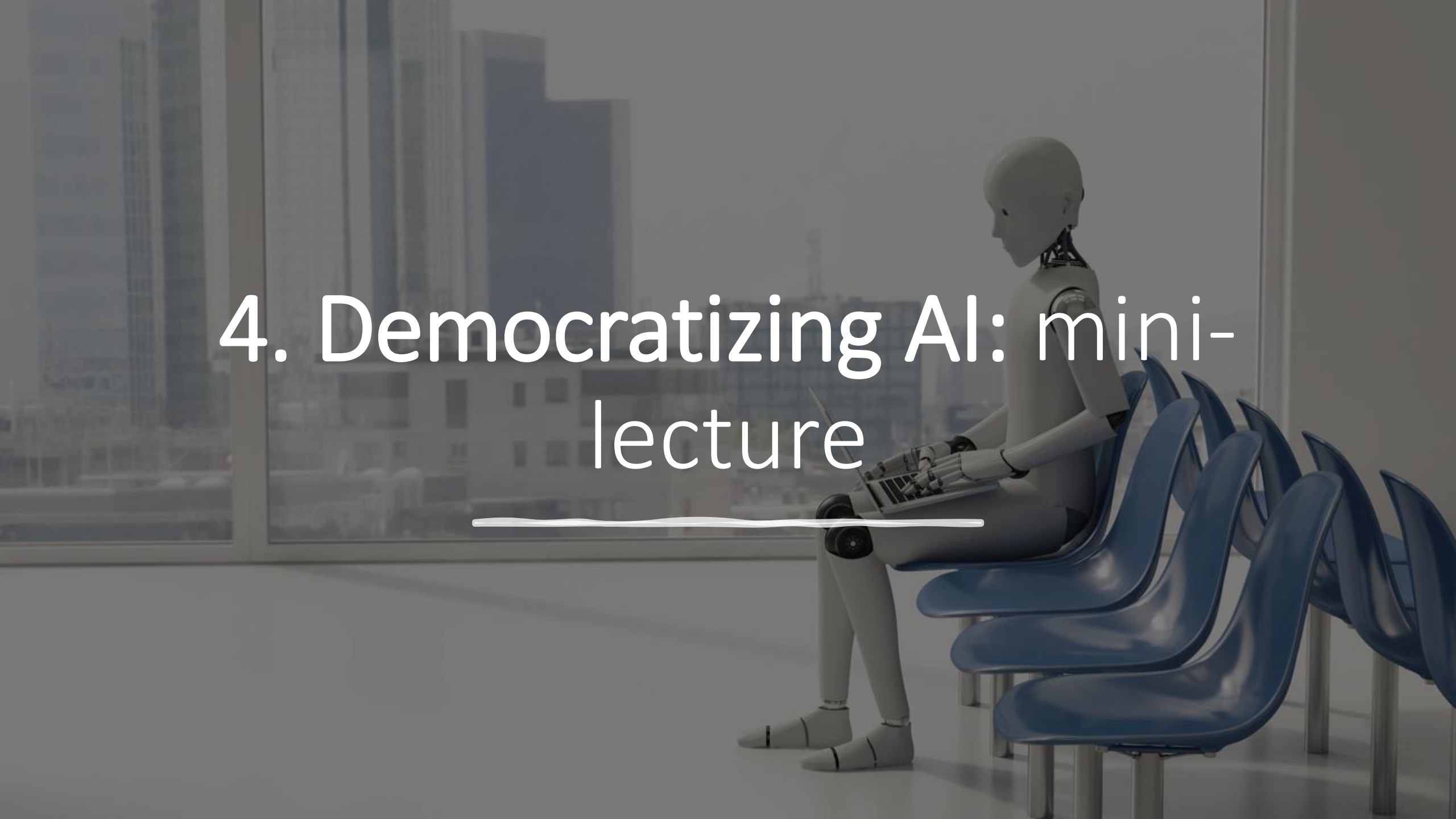
(Weak) AI and Democracy 1/2

1. What is the impact of AI on democratic institutions?
2. Should AI itself be *democratized*?
   - Do "explicability" and accountability require democratization?
   - *Normative/ethical question: Is democratization of AI desirable?*
   - *Practical Question: What would this look like in practice?*

(Weak) AI and Democracy 2/2

- What is the value of democratic institutions?
  - Autonomy: democratic processes aim at self-governance & freedom for citizens
  - Justice: democratic processes aim at fair distribution of social benefits and burdens (fairness may be an outcome of or secured by democratic deliberation)
  - Stability: if participatory democracies are fairer, and if fairness promotes stability, then democratic societies should be more stable than alternative arrangements (all else being equal).
- How can AIs impact democratic institutions?
  - Democracy depends on free and informed citizens. AI in journalism, social media engagement algorithms, algorithms behind search-engine recommendation, etc., shape the information space and have the power to create division or cohesion, to promote or suppress pro-democracy activism (e.g., YouTube radicalization, Facebook and violent extremism), and more
  - AI integrated into democratic decision-making processes may call into question the legitimacy of democratic institutions (Beckman et al, 2022)
  - AI arms race with authoritarian nations with different views of (e.g.) human dignity (and thus privacy, freedom of speech and assembly, etc.)
  - The threat of autonomous AIs (The “Superintelligence” scenario)

Beckman et al. 2022. “Artificial intelligence and democratic legitimacy. The problem of publicity in public authority” *AI ad Society*



4. Democratizing AI: mini-lecture

What might it mean to democratize AI?

To make AI technologies broadly available and accessible to everyone or give everyone the resources to use/develop AI tech for themselves

To ensure that AI technologies are governed democratically.

Reasons to "democratize" AI

*"…the **fundamental democratic norm** where all decisions (say of parliament or public officials) are documented and open to inspection by citizens at a later time. Similarly, this would be translated to public sector AI systems, and more generally, any technology or AI system that claims to be in the public interest. It speaks to the idea that democratic civil societies not only have the right to understand the workings of technology (which requires transparency and explainability) but also be able to validate its mechanisms to be democratic if they are claimed to serve the public interest."(Züger & Asghari 2022, emphasis added)*

"AI for the public. How public interest theory shifts the discourse on AI" (Züger & Asghari 2022)

Reasons to democratize AI continued

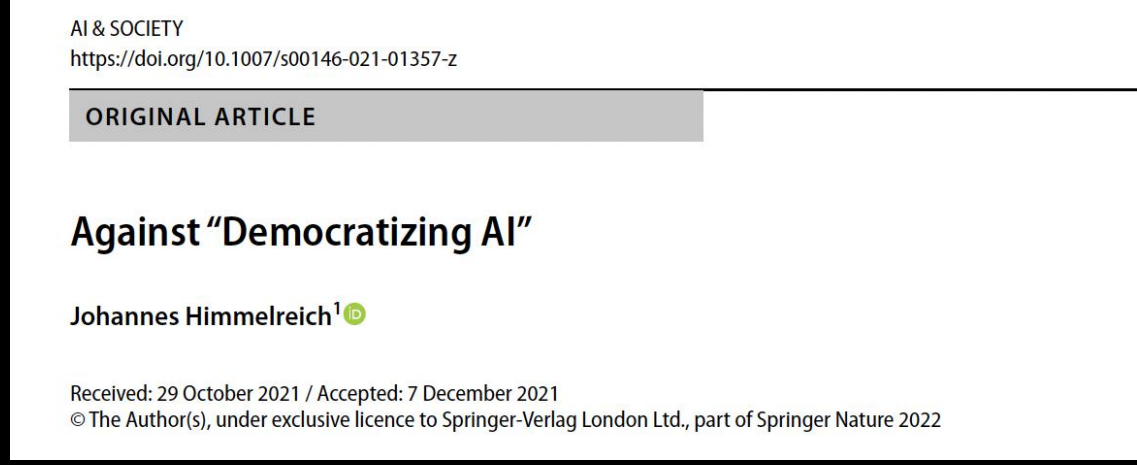
Rationale 1: AI democratization as an extension of democratic principles of self-governance by the people. Insofar as AI is both a public good and a constitutive element of collective democratic practices, it should be under popular control.

> *"... democratic civil societies not only have the right to understand the workings of technology (which requires transparency and explainability) but also be able to validate its mechanisms to be democratic if they are claimed to serve the public interest."(Züger & Asghari 2022; emphasis added)*

Rationale 2: Democratizing AI is a strategy for addressing existing injustices (e.g., in policing, education, etc.) and ensuring other moral goods, such as equality, fairness, and freedom.

> *"AI for the public. How public interest theory shifts the discourse on AI" (Züger & Asghari 2022)*

Limitations of the standard approaches to AI democratization



AI & SOCIETY
https://doi.org/10.1007/s00146-021-01357-z

ORIGINAL ARTICLE

Against "Democratizing AI"

Johannes Himmelreich[1]

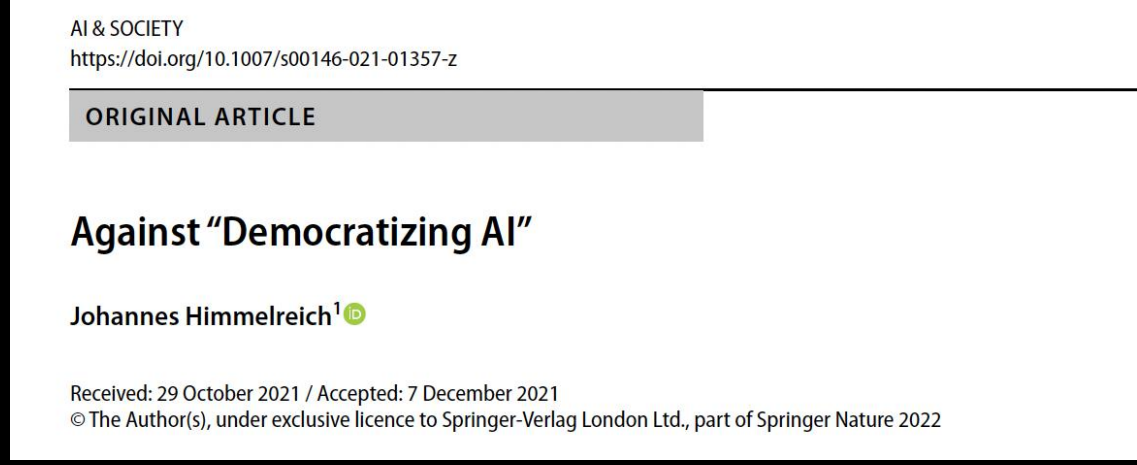
Received: 29 October 2021 / Accepted: 7 December 2021

"Against Democratizing AI" (Himmelreich 2021): *"Democracy is a system of social governance that gives everyone a fair and equal opportunity of influence."*

Proposals to democratize AI typically construe democratization as broad and maximally inclusive "participation by the people."

- Too abstract – does not specify *how* participation should happen – and thus open to exploitation.
- Language of democracy (broad participation by the people) has been co-opted by authoritarians, kleptocrats, and populists.

Limitations of the standard approaches to AI democratization



AI & SOCIETY
https://doi.org/10.1007/s00146-021-01357-z

ORIGINAL ARTICLE

Against "Democratizing AI"

Johannes Himmelreich[1]

Received: 29 October 2021 / Accepted: 7 December 2021

"Against Democratizing AI" (Himmelreich 2021):

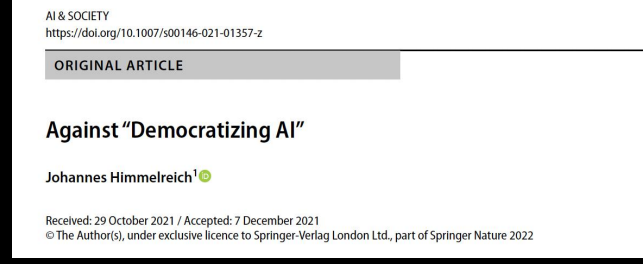
Critiques standard calls for AI democratization + proposes that AI should be democratized not by aiming at increased participation *simpliciter* but by "democratizing the administrative and executive components of our collective decision-making."

Recommends asking 3 questions:

1. The democratization of *what?*
2. *Why* should AI be democratized?
3. *How* should AI be democratized?

Himmelreich's 5 reasons against Democratizing AI

1. Weak grounds
2. Redundancy
3. Resource intensity
4. Popular oversights
5. Theoretical and practical inefficacy



AI & SOCIETY
https://doi.org/10.1007/s00146-021-01357-z

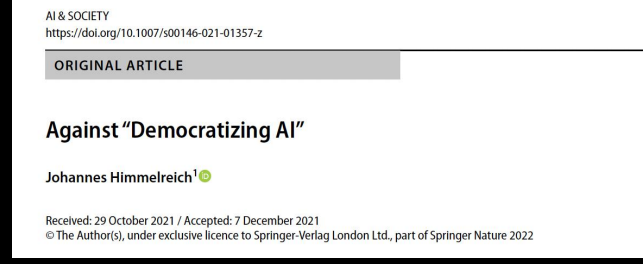
ORIGINAL ARTICLE

Against "Democratizing AI"

Johannes Himmelreich[1]

Received: 29 October 2021 / Accepted: 7 December 2021

Himmelreich's 5 reasons against Democratizing AI 1/5



AI & SOCIETY
https://doi.org/10.1007/s00146-021-01357-z

ORIGINAL ARTICLE

Against "Democratizing AI"

Johannes Himmelreich[1]

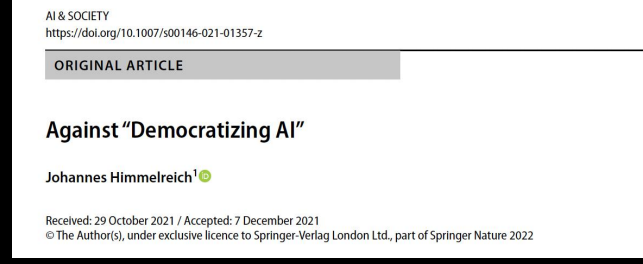
Received: 29 October 2021 / Accepted: 7 December 2021

1. Weak grounds –AI systems are not the kinds of thing that need to be legitimized by being democratized, and democratization may exacerbate the very problems that it is meant to address.

"Whereas democracy might well be justified as the governance system of a state, and while it may contribute to justifying state authority, democracy might not be justified as a governance system beyond the state."

2. Redundancy
3. Resource intensity
4. Popular oversights
5. Theoretical and practical inefficacy

Himmelreich's 5 reasons against Democratizing AI 1/5



AI & SOCIETY
https://doi.org/10.1007/s00146-021-01357-z

ORIGINAL ARTICLE

Against "Democratizing AI"

Johannes Himmelreich[1]

Received: 29 October 2021 / Accepted: 7 December 2021

1. Weak grounds

Triggers of legitimacy (apply to the state, but not strongly enough to AI):

i. Coercive powers – AI exists in already coercive environments; may contribute to existing coercive schemes, but not produce novel systems of coercion

ii. Pervasive impact – AI contributes to systems that have pervasive impact, but does not create novel ones

iii. Involvement in a scheme of social cooperation – "*AI is embedded in existing systems of social coordination, but it does not constitute a system of social coordination.*"

Democratization has costs (resource intensive, decisions by committee can lead to suboptimal outcomes and thus frustrate the aims of fairness)

2. Redundancy

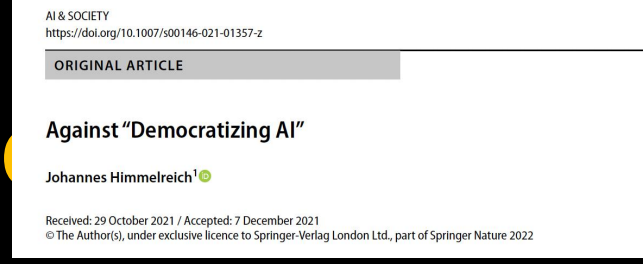
3. Resource intensity

4. Popular oversights

5. Theoretical and practical inefficacy

Himmelreich's 5 reasons against Dem
AI 2/5

1. Weak grounds
2. Redundancy – *"In domains, where the state governs legitimately through democratic institutions, no other institution ought to compete over practical authority." Insofar as oversight/regulatory institutions already exist, novel ones are unnecessary.*
3. Resource intensity
4. Popular oversights
5. Theoretical and practical inefficacy


AI & SOCIETY
https://doi.org/10.1007/s00146-021-01357-z

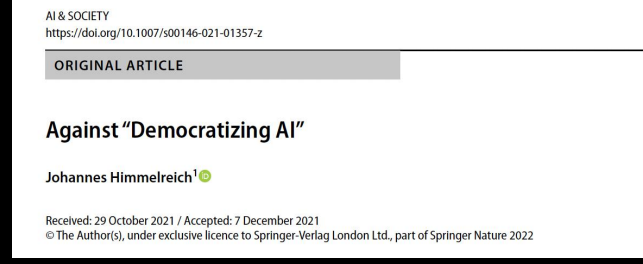
ORIGINAL ARTICLE

Against "Democratizing AI"

Johannes Himmelreich[1]

Received: 29 October 2021 / Accepted: 7 December 2021

Himmelreich's 5 reasons against Democratizing AI 3/5



AI & SOCIETY
https://doi.org/10.1007/s00146-021-01357-z

ORIGINAL ARTICLE

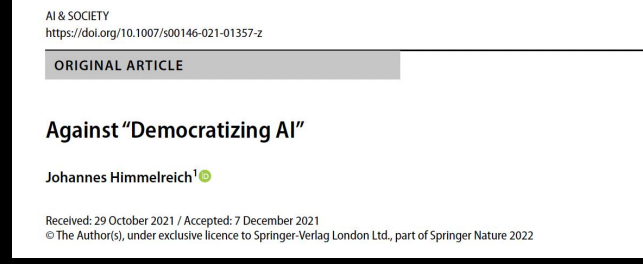
Against "Democratizing AI"

Johannes Himmelreich[1]

Received: 29 October 2021 / Accepted: 7 December 2021

1. Weak grounds
2. Redundancy
3. Resource intensity – implementational challenges due to high demand for cognitive, financial, & temporal, resources of democracy—and especially direct democracy--can lead to democratization of AI being "infeasible, inefficient, or inequitable." Compounding factor: the least advantaged may have the fewest of these recourses, so would be marginalized further. This leads to tension between participation & deliberation.
4. Popular oversights
5. Theoretical and practical inefficacy

Himmelreich's 5 reasons against democratizing AI 4/5



AI & SOCIETY
https://doi.org/10.1007/s00146-021-01357-z

ORIGINAL ARTICLE

Against "Democratizing AI"

Johannes Himmelreich[1]

Received: 29 October 2021 / Accepted: 7 December 2021

1. Weak grounds
2. Redundancy
3. Resource intensity
4. Popular oversights – call to democratize AI emphasizes participation over deliberation. But, without systems to ensure that participation yields rational and unbiased collective decision-making, the process is susceptible to cognitive biases, motivated reasoning, groupishness, and may result in moral short-sightedness.
5. Theoretical and practical inefficacy

Himmelreich's 5 reasons against Democratizing AI 5/5

1. Weak grounds
2. Redundancy
3. Resource intensity
4. Popular oversights
5. Theoretical and practical inefficacy – *"A final reason against the call to democratize AI is not that there is something wrong with it, but that the call to democratize AI fails to speak to the concerns that animate it [injustice, oppression, etc.]"* since *"[d]emocracy is normatively thin." "deliberative conception of democracy" is a step is preferable, but also insufficiently thick.*

Conclusion

Rather than emphasizing popular participation – the broader and deeper the better – emphasis should be on shoring up democratic processes within existing governance structures.

Exercises

Take a minute to evaluate these objections to the standard picture.

Is there a practical dimension to the proposal that we make existing structures either more democratic or more responsive to the reasons/causes that animate the call for democratizing AI?

What might your role be in evaluating or implementing these recommendations? As AI researchers, developers, citizens, etc?

5 minute BREAK

6. Can AIs participate in democratic decision-making?

First, an ambiguity

Q: How, if at all, should AIs participate in democracy?

A: Depends on whether we're talking about weak AI or strong AI/AGI.

The weak-AI participation in democratic practices: As support for machine learning systems and algorithms supporting democratic procedures (e.g., in voting systems, systems of public information exchange, including digital and social media algorithms, etc.)

The Strong AI/AGI participation: As any autonomous agent of the polis with rights consistent with citizenship?

“The Democratic Inclusion of Artificial Intelligence?”

From moral patiency/moral agency to political patience/political agency

The Democratic Inclusion of Artificial Intelligence? Exploring the Patiency, Agency and Relational Conditions for Demos Membership

Ludvig Beckman[1] · Jonas Hultin Rosenberg[2]

Received: 19 October 2021 / Accepted: 12 March 2022 / Published online: 30 March 2022

Species-ism and the democratic inclusion of AI (AGI/S-AI)

- Objection out of the gates: AI are not human, so how can they have the same rights to political participation as humans?
- Response: Assume that speciesism is wrong (that all and only humans can qualify for democratic participation just because they are humans).
- Question to be addressed: what *properties* ground right to democratic participation

"Our suggestion is that the properties that future AI needs to possess in order to qualify for democratic inclusion are the properties that an entity needs to possess in order to have political agency and political patiency—i.e. agency and patiency in the sense relevant for democratic inclusion."

Some issues to be addressed

- The (moral and political) rights of technological versus biological beings.
- Whether AIs, as designed artefacts, can have intentions separate from those of the designer.
- What is required for AIs to be autonomous in the sense required for political participation.
- Whether an AGI needs to have affect as well as cognition in order to count morally and be represented politically.
- Whether AGIs can experience not just intelligence (cognition) but affect (sensation/emotions)?

Even if S-AI/AGI have moral and legal rights—because it is conscious, autonomous, and intelligent—do these rights extend to the right to political participation?

Child android Affetto reacting to tactile inputs

Can AIs bear the right kind of relationship to the state?

Two competing criteria for democratic inclusion:

AAP: "...an entity is entitled to inclusion in the demos if and only if [it is] *affected by* the decisions of the political unit in the relevant sense."

ASP: "…an entity is entitled to inclusion in the demos if and only if [it is] *subjected to* the decisions of the political unit in the relevant sense."

Can AIs bear the right kind of relationship to the state?

AAP: "...an entity is entitled to inclusion in the demos if and only if [it is] *affected by* the decisions of the political unit in the relevant sense."

- Stretches boundary of inclusion to everyone that have interests that can be causally affected by the decisions of the political entity. May includes children, animals, and AI provided they have relevant interests.
- Which interests are relevant & does AI have them?
  - AI plausibly has "teleo interests" (Basl), but this seems too expansive (even bacteria have these)
  - Limit to interests that require a "sense of self-hood and purposive action"?

Can AIs bear the right kind of relationship to the state?

ASP: "…an entity is entitled to inclusion in the demos if and only if [it is] *subjected to* the decisions of the political unit in the relevant sense."

Subjection understood as subjection to law, not just subjection to coercion. Two interpretations:

1. *The first holds that an entity is relevantly subjected to public decisions if and only if the entity is legally obligated to comply with the decision.*
   - *AI can be subject to legal duties only they are the kind of entity that can be the bearer of legal duties (i.e., a legal person) and if it can recognize and respond to rules (i.e., it is the capacity for action)*
2. *The second holds that an entity is relevantly subjected to public decision if and only if the entity is subject to claims to legitimate authority.*
   - *Requires agency as well as the capacity to recognize the legitimacy of the law. More stringent than the first.*

Does AI have agency in the sense necessary for inclusion on the ASP?

- *"the first reading of the ASP holds that an entity is subjected to the law in the relevant sense if and only if it is a legal person within the jurisdiction of the legal system: the law applies to legal persons only."*

- *"the second reading of the ASP, an entity is subjected to the law in the relevant sense if and only if it is a legal person within the jurisdiction of the legal system that possesses the ability to comply with the law and to recognize the law's claim to legitimate authority."*

Do AIs need to be proper agents for the purposes of democratic inclusion?

- Legal personhood - current and future AIs may meet these conditions
- Compliance - not all entities that are granted legal personhood (e.g., ecosystems) are capable of complying with the law.
  - Skepticism: the law is a moving target and not explicitly stated, thus pre-programmed rules won't work and rigid approaches to problem-solving may similarly fail.
  - Yet: AI may approach levels of human intuition
- Recognition of Appropriate Authority
  - Does AI have the capacity to recognize the law as legitimate?
    - Norm-recognition capacity is necessary but not sufficient
    - Capacity to recognize legitimate authority is a moral capacity

Upshot for ASP

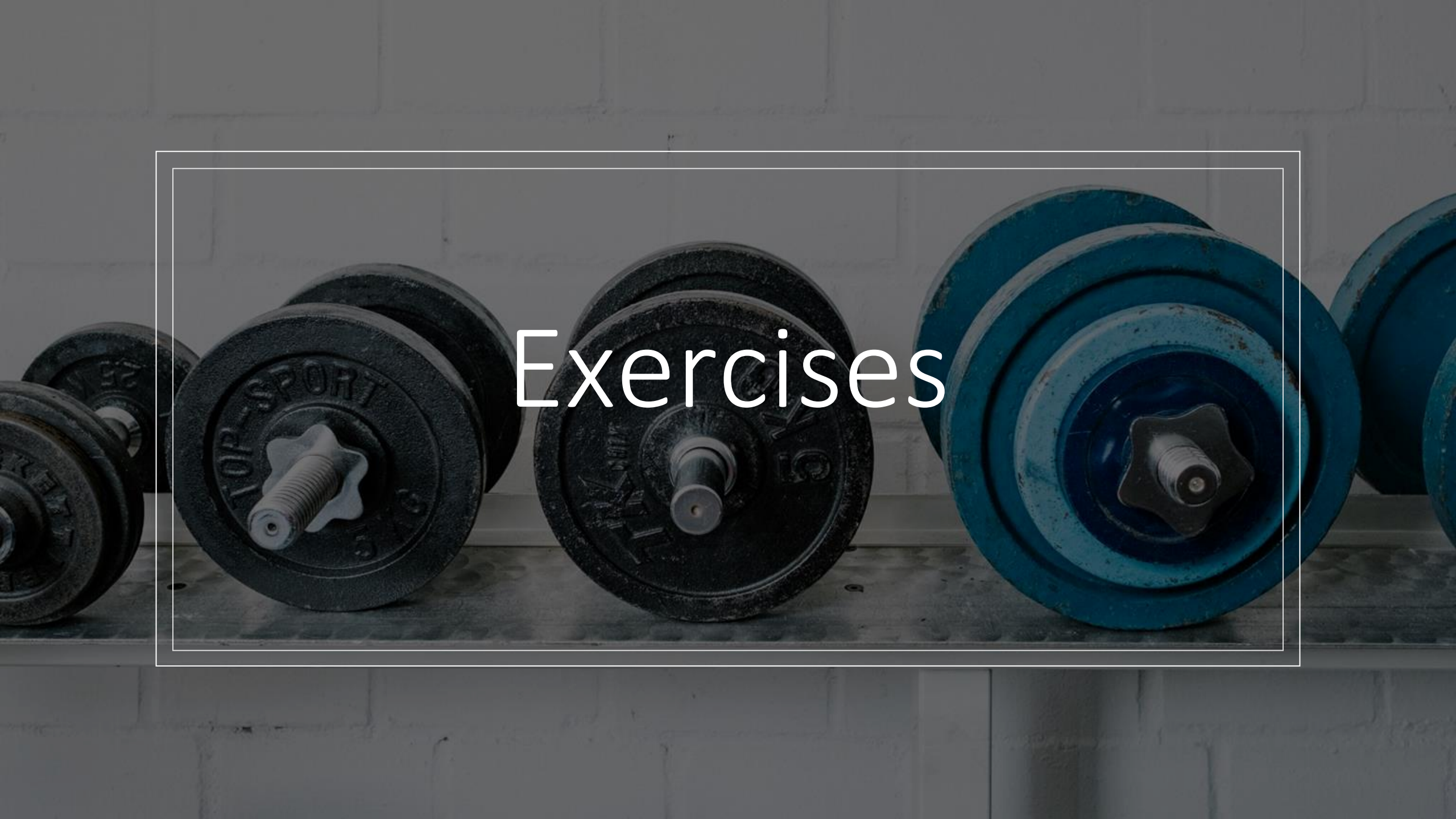
"With respect to AI, the implication is that they should be included in the demos only if they have the ability to make moral judgments about the legitimacy of legal authority."

Moral and political patiency

- AAP: Political patiency = an entity has interests that warrant its political inclusion (not the same as having interests that deserve political protection)
  - Which interests matter? Psychological interests (not merely teleo interests)
  - Autonomy
    - Kantian view collapses moral agency with moral patiency
    - Autonomy as life plans (but implications for children and severely cognitively limited adults)
  - Consciousness (as subjective experience)
    - Inclusion limited at sentience, but broadly inclusive

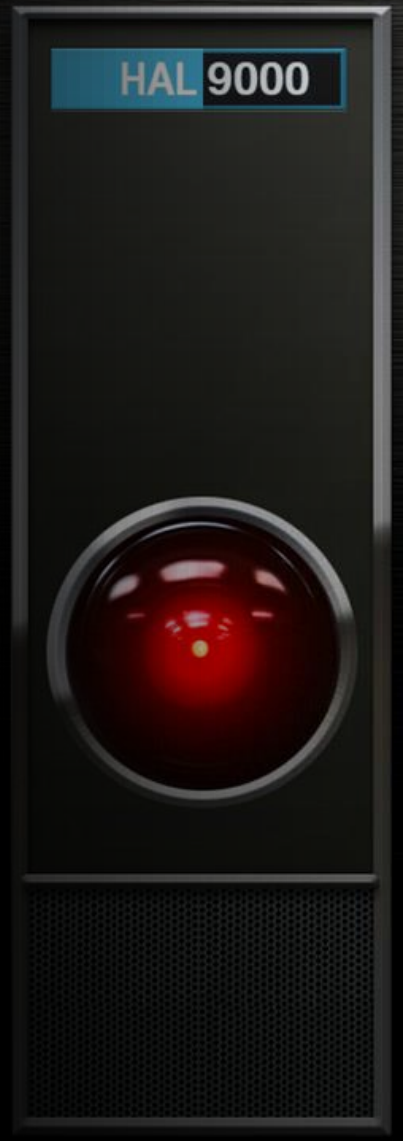
Summary

“AI’s can be either affected or subjected to collective decisions. But … it is less clear that AI’s do or can satisfy agency and patiency requirements. Democratic inclusion cannot be conclusively determined from the fact that an entity is either subjected or affected by public decisions: only agents and/or patients qualify as members of the demos.”



Exercises

Thank you!



Here are some extras, Dave.

5. Case Studies in Ethical AI

Exercises through the lens of the five principles

Applying the Five Bioethical Principles further

Floridi et al. focus on making recommendations based on the 5 principles to governments, but what about other stakeholders? Turn to case studies.

As we consider the case study questions:

… consider how your answers draw on the five bioethical principles:

Beneficence – *non-maleficence* – *autonomy* – *justice* – *explicability*

Case Study A:
Law Enforcement Chatbots

Case Studies are the product of the Princeton University Center for Human Values (UCHV) and the Center for Information Technology Policy (CITP) at Princeton. **http://www.aiethics.princeton.edu**

Case Study B : *Optimizing Schools*

Case Studies are the product of the Princeton University Center for Human Values (UCHV) and the Center for Information Technology Policy (CITP) at Princeton. **http://www.aiethics.princeton.edu**

Case Study C: *Automated Healthcare App*

Case Studies are the product of the Princeton University Center for Human Values (UCHV) and the Center for Information Technology Policy (CITP) at Princeton. **http://www.aiethics.princeton.edu**