

# Human Dignity and Artificial Intelligence

Mapping the Contours and Utility of Human Dignity in addressing AI Challenges

Sue Anne Teo  
Center for European and Comparative Legal Research (CECS), Faculty of Law,  
University of Copenhagen

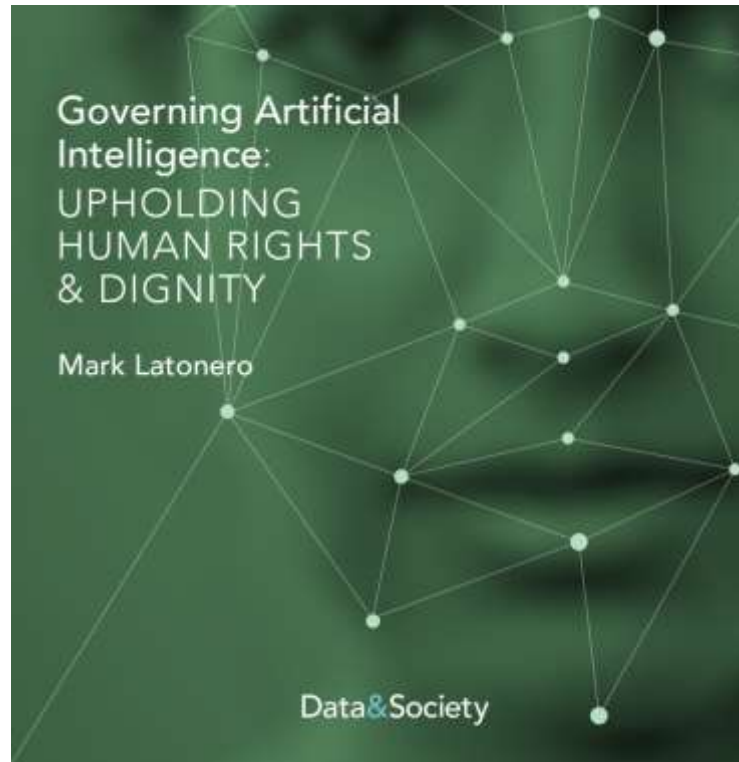
KØBENHAVNS UNIVERSITET



# Outline

- Motivation
- A short history of human dignity
- Human dignity within human rights law
- Conceptions of human dignity
- Human dignity and AI: Connecting the dots
- Human dignity and AI: Plotting new dots

# Motivation



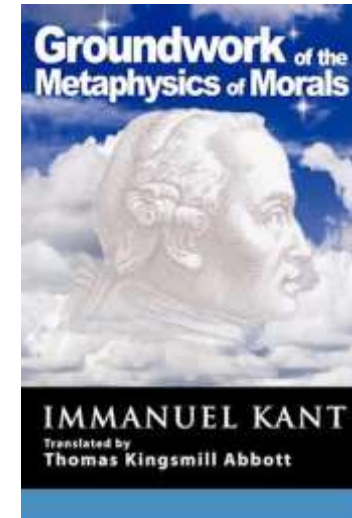
In order for AI to benefit the common good, at the very least its design and deployment should avoid harms to fundamental human values. International human rights provide a robust and global formulation of those values.

In bridging AI and human rights, what's at stake is human dignity.\* As an international framework, human rights law is intended to establish global principles ("norms") and mechanisms of accountability for the treatment of individuals. As such, a rights-based

\* The definition of human dignity is contested and its normative value is debated in an extensive literature that is outside the scope of this report. For the present purposes, the term human dignity gestures towards its usage in Western moral philosophy, such as Kant's notions of dignity linked to human autonomy and agency, while acknowledging that dignity has been linked to traditions such as Eastern philosophy as well. This report's usage of human dignity also evokes the United Nation's charter, the Universal Declaration on Human Rights, and the major rights treaties, which link fundamental human rights, the dignity and worth of the human person, and the equal rights of men and women. The interactions between humans and AI may further challenge or refine the concept of human dignity, which is an important topic for future work.

# A short history of human dignity

- Dignity as a status concept
  - Roman antiquity
  - Holder of office, position, rank
  - Conduct befitting that position
- Dignity as inherent worth
  - By virtue of being human
  - Cicero: distinct from non-human animals due to our capacity for reasoning
  - Aquinas: inherent worth in the human person, in line with divine will
  - Immanuel Kant: formula of humanity and the autonomy of humans as moral actors



## Groundwork on the Metaphysics of Morals (1785)

» “In the kingdom of ends everything has either price or dignity. Whatever has a price can be replaced by something else which is equivalent; whatever, on the other hand, is above all prices, and therefore admits of no equivalent, has a dignity.

Whatever has reference to the general inclinations and wants of mankind has a market price; whatever, without presupposing a want, corresponds to a certain taste, that is to a satisfaction in the mere purposeless play of our faculties, has a fancy price; but that which constitutes the condition under which alone anything can be an end in itself, this has not merely a relative worth, I.e., price, but an inner worth, that is dignity.”

(Ak. 4:434-435)

= Dignity as an absolute value

## Groundwork on the Metaphysics of Morals (1785)

“Now morality is the condition under which alone **a rational being can be an end in himself**, since by this alone is it possible that he should be a legislating member in the kingdom of ends. Thus morality, and humanity as capable of it, is that which alone has dignity.”

(Ak. 4:434-435)

Dignity = inhere in humans as moral actors

# Human Dignity within International Human Rights Law

- Said to play a foundational role
- UDHR
- Derivative – language of ICCPR and ICESCR
- Recognition of an ‘intrinsic worth’
- Three-pronged role in human rights law:
  - Foundational
  - Standard to reach for in relation to other rights
  - A stand-alone right (eg. Article 1 CFR)



# Conceptions of a concept

- Human dignity lacks a clear definition
- McCrudden: 'minimum core content'
  - Intrinsic worth by virtue of being human
  - This intrinsic worth should be recognised and respected by others
  - State under an obligation to protect human dignity
- Conceptions approach taken instead
- W.B.Gallie 'essentially contested concepts'
- Point of departure based on the fact that neither a consensus on what the concept means or consistency in its use exists: McCrudden; Rosen





# The 4 conceptions of human dignity

1. Non-instrumentalization of persons
2. Protection of vulnerable classes of persons
3. Recognition and expression of self-worth
  - Autonomy
  - Physical and mental mistreatment
  - Enabling material conditions
4. Protection of humanity as a species concern



# Human Dignity and AI: Connecting the Dots

- No need to reinvent the wheel
- New circumstances, new technologies can be accommodated within the existing conceptions of human dignity
- Definition of AI
- Examples:
  - Lethal autonomous weapons systems – engages ‘protection of humanity as a species concern’ and ‘non-instrumentalization of human persons’
  - Bias and discrimination within AI systems – engages the ‘non-instrumentalization’ and the ‘autonomy’ conception
  - Platform content moderation and curation by AI systems – engages the ‘recognition and expression of self-worth’ conception of human dignity, notably through respect for informational privacy

# Human Dignity and AI: Plotting New Dots

- No need to reinvent the wheel but there is a need to address some novel challenges brought forth by AI systems
1. The lived experiences of persons subjected to AI systems should inform the contours of the meaning of human dignity in the age of AI.
    - Human dignity as embodied lived experiences, not abstract philosophical idea
    - New affordances – harms can result in novel, unexpected ways; lack transparency; relationships of power
    - Harms inhere in systems – creating precarity of background conditions, not obvious, not observable, cannot pass through the threshold of it amounting to harm
    - Dignity-as-fairness

# Human Dignity and AI: Plotting New Dots

## 2. Disembodiment of empiric self-representation and contextual sense-making

- Separation of self-representation and the data self: an unnecessary condition – data is relational, correlations and inferences
- Community informed normative standards are displaced through algorithmic standards – eg. Algorithmic management of workers – displace contextualised standards – replaced with algorithmically optimised standards
- Dignity-as-disembodiment

# Human Dignity and AI: Plotting New Dots

## 3. Cognitive autonomy through adaptive choice architectures

- Material affordance of AI
- Failure to operationalise the protections for freedom of thought and freedom of opinion (rights vernacular is concerned with manifestations of these freedoms) – issue of substance
- Indirect challenges such as through adaptive choice architectures as condition antecedent for cognitive autonomy (*forum internum*) remains under-explored – issue of form
- Dignity-as-contextual affordance

# Conclusion

- No need to reinvent the wheel as the four conceptions go a long way
- However, in order to respect and protect human dignity in the age of AI, three key aspects must be foregrounded:
  - Embodied lived experiences (dignity-as-unfairness)
  - Disembodiment of empiric self-representation and contextual sense-making (dignity-as-disembodiment)
  - Cognitive autonomy through choice architectures (dignity-as-contextual affordance)

Thank you for your attention

Comments and questions welcomed @ [sue.anne.teo@jur.ku.dk](mailto:sue.anne.teo@jur.ku.dk)

