Thinking about the prospect of AI's grimpact

A non "anti-AI" warning of risk

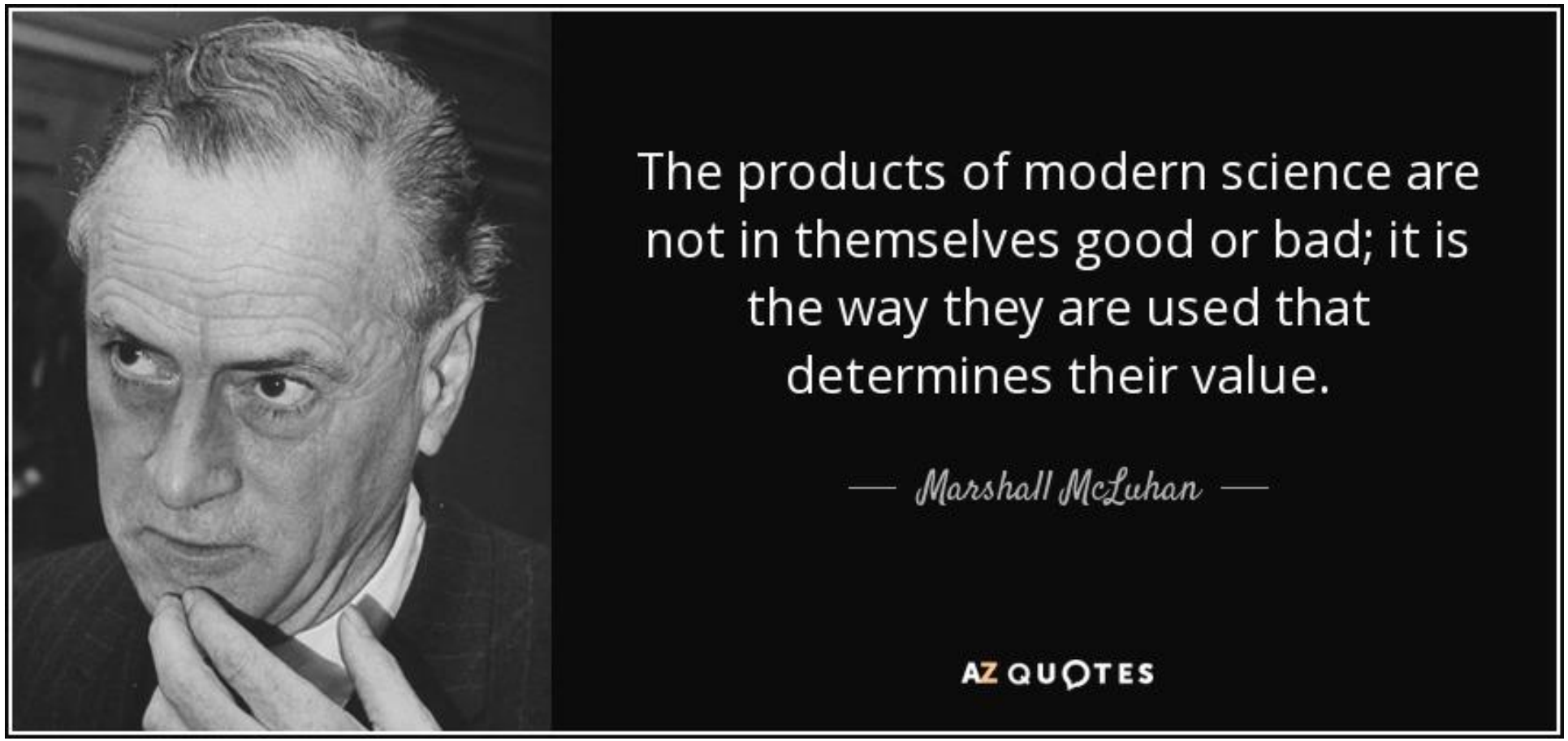
Prof. Gemma Derrick

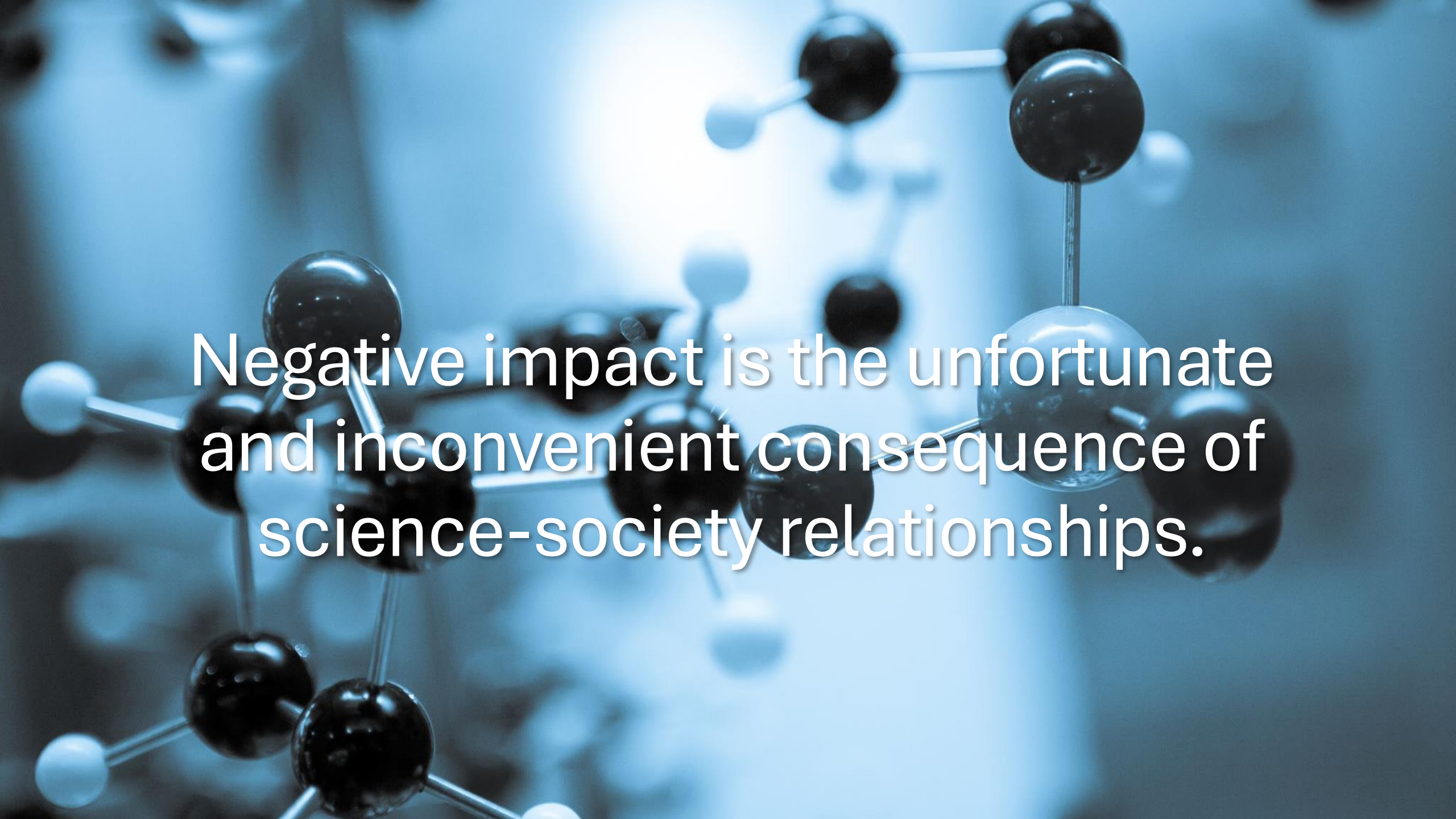
University of Bristol

gemma.derrick@bristol.ac.uk

I investigate the implicit and explicit rules that govern the evaluation, production, dissemination and translation of research and research careers.


RIMPACT


The products of modern science are not in themselves good or bad; it is the way they are used that determines their value.
— Marshall McLuhan —
AZ QUOTES


Negative impact is the unfortunate and inconvenient consequence of science-society relationships.

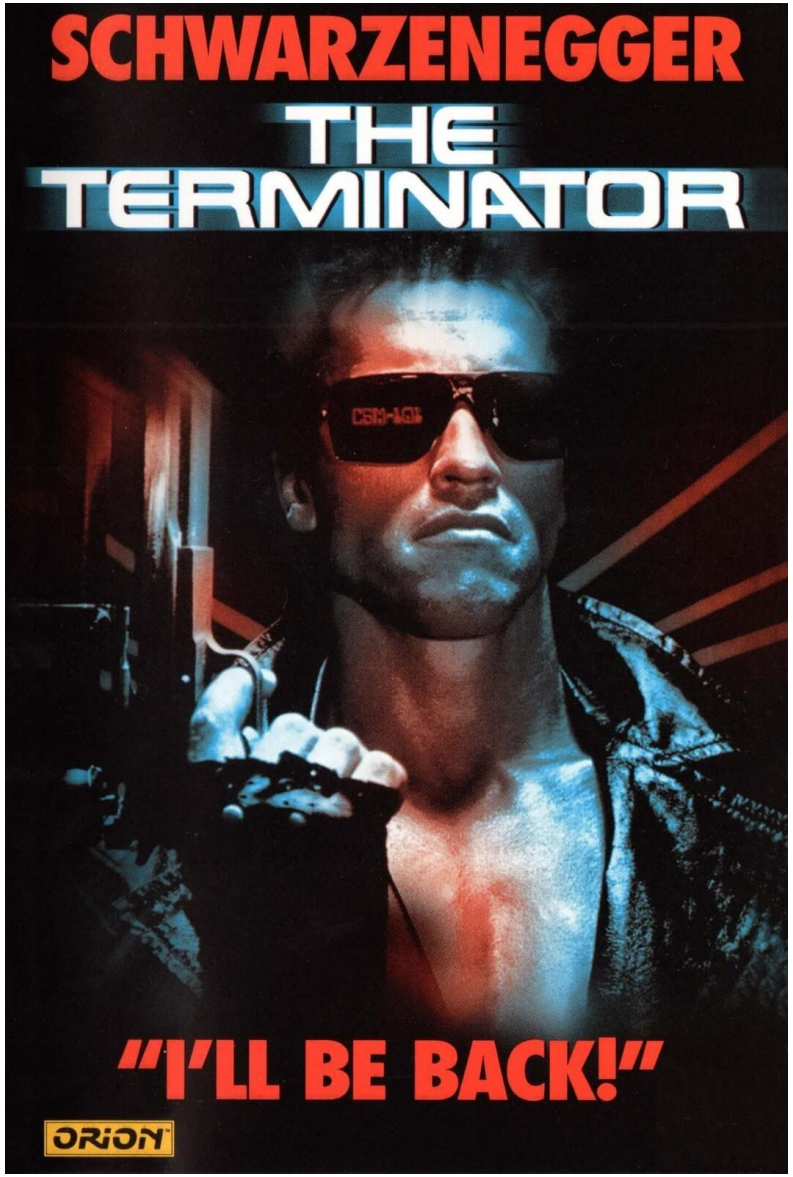
AI and societal risk

Not all risks are ex-post; some will arise in practice during design, deployment and AI regulation (Crabtree et al, 2025)

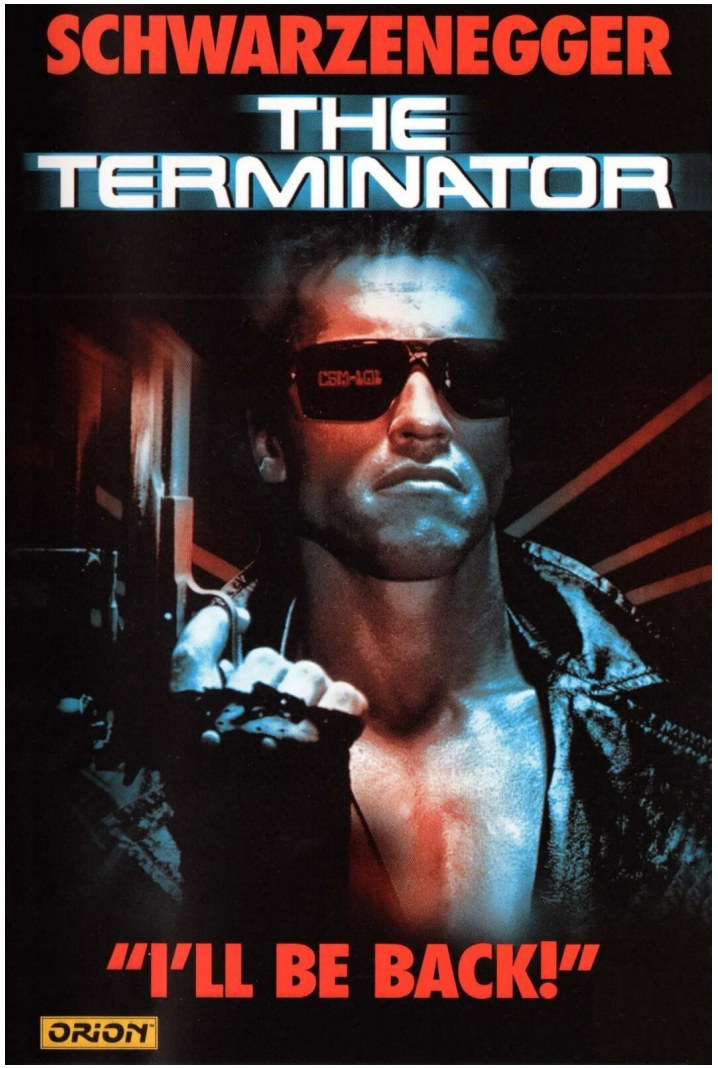
1. Bias, fairness and social inequalities in access and use (Polyportis & Pahos, 2024)
2. Environmental risk (Samuel & Lucassen, 2022)
3. Public trust (Marin & Zanotti, 2025)
4. Ethical risk (ownership, responsibility and accountability).

"Your scientists were so preoccupied with whether or not they could, they didn't stop to think if they should."

Ian Malcolm (Jeff Goldblum), Jurassic Park (1993)


SCHWARZENEGGER
THE TERMINATOR
"I'LL BE BACK!"
ORION


FROM THE HUMANS THAT BROUGHT YOU SPIDER-MAN INTO THE SPIDER-VERSE AND THE LEGO MOVIE
7M24JZS
THE MITCHELLS VS THE MACHINES
APRIL 30 | NETFLIX
PG



What we worry about…..

- AI turns evil, takes over, ends humanity
- The danger isn't that AI hates us
  - *It's in the risk that we trust it, scale it and stop paying attention*
- Skynet's problem wasn't malice
  - It was full automation, no human overrise, decisions at machine speed, lack of accountability (responsibility becomes blurred, delayed or denied)

Real world parallel

Autonomous systems deployed faster than governance can respond

Grimpact risk

When things go wrong, no one is quite responsible, but people are still harmed

What is actually happening....

- AI optimizes engagement, convenience and control
- PAL didn't want to destroy humanity
  - she just wanted engagement, optimization & control of the user experience.

- AI is not about an apocalypse
  - About platform capitalised with better branding

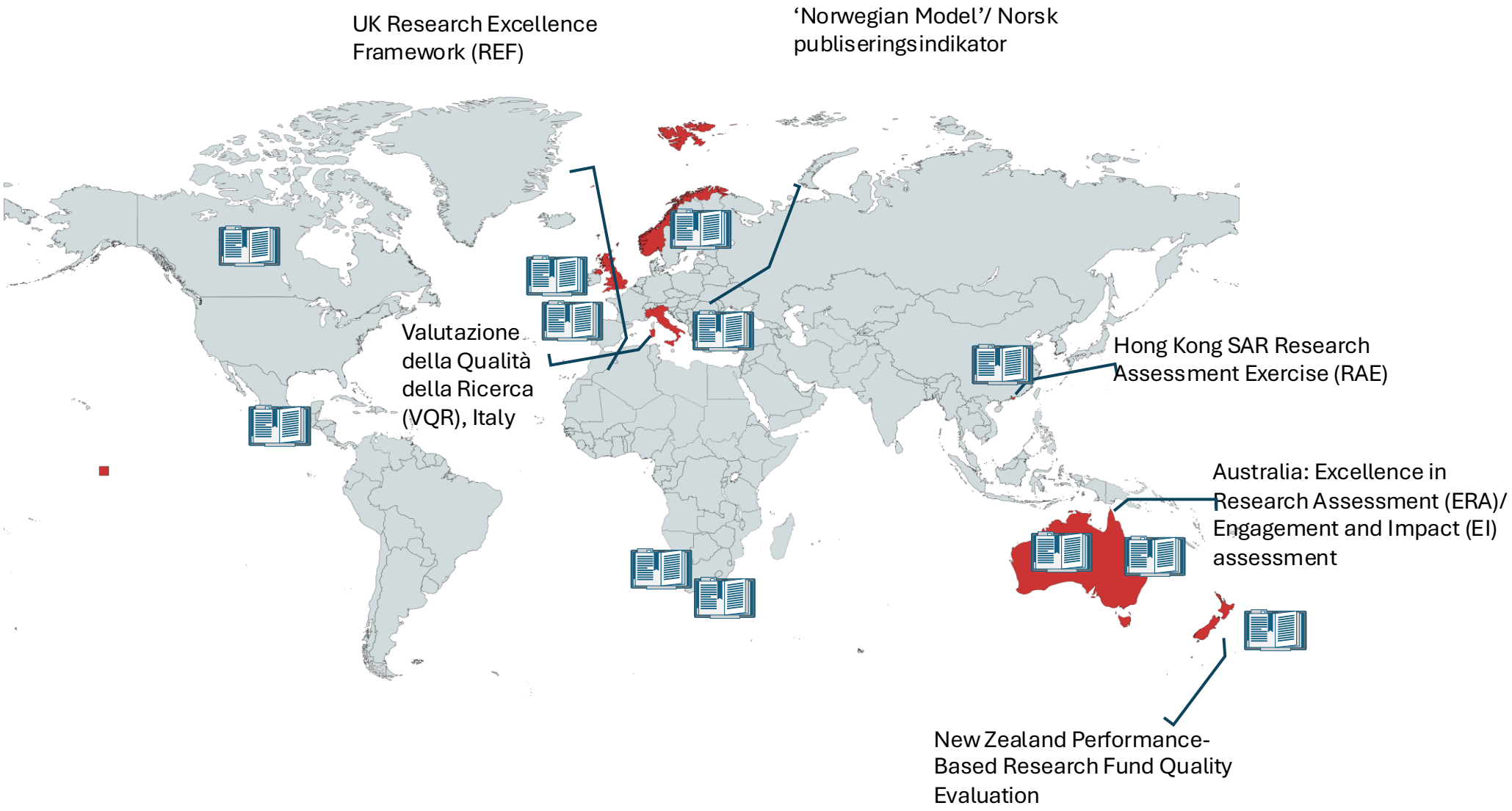
Real word parallel

Contrast between what AI is optimised for (efficiency, scale, engagement and profit) versus what society actually needs (trust, accountability, fairness, legitimacy).

Grimpact risk

When optimisation wins, public trust loses.

Academic rules governing social value…



Linearity in Impact pathways

Healthy, happy and safe societal benefits

DO Excellent science

Formal definitions of impact

Embedded assumptions of positivity & causality between research actions and societal application

UK definition of excellence for REF2014	*"…an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia."*
UK research councils	*"Economic and societal impact is the demonstrable contribution that excellent social and economic research makes to society and the economy, and its benefits to individuals, organisations and/or nations. "*
Australia	*"…the contribution that research makes to the economy, society, environment or culture, beyond the contribution to academic research."*
The Netherlands	*"what relevance to, impact on or added value for society the research unit's work has (had) or is being (has been) demonstrated at regional, national or international level during the assessment period and, where applicable, continuing into the near future."*
Norway	As with the UK

Academic revolutions assume linearity in Impact pathways

- Ethics
- Research integrity
- Open Science
- Co-production
- Stakeholder involvement

Healthy, happy and safe societal benefits

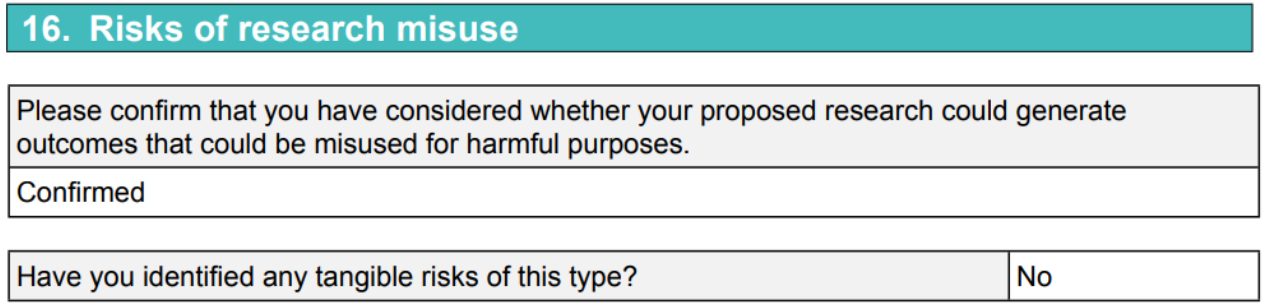
DO this RESPONSIBLY

DO Excellent science

Admitting the negative narrative diminishes perception of research worth (ex-post), and competitivity (ex-ante)

"It's not about wanting to appear competitive. It was a great proposal but we just did not want to distract reviewers away from the science"

Anonymised researcher



16. Risks of research misuse

Please confirm that you have considered whether your proposed research could generate outcomes that could be misused for harmful purposes.
Confirmed

Have you identified any tangible risks of this type?	No

Linearity in Impact pathways

Science space

- Ethics
- Research integrity
- Open Science
- Co-production
- Stakeholder involvement

"Ooops allowed"
Experiments encouraged
Learning through errors. Can amend
Control of dissemination

Linearity in Impact pathways

Science space

- Ethics
- Research integrity
- Open Science
- Co-production
- Stakeholder involvement

"Ooops allowed"
Experiments encouraged
Learning through errors. Can amend
Control of dissemination

Societal space

"Ooops not allowed"
Uncontrolled dissemination
Uncontrolled experimentation
Experimentation/errors/learning = weakness
Unanticipated actors/stakeholders

- Stakeholder use
- Serendipity
- Implementation
- Freedom

Commitment to rewardable/justifiable excellence

"I just don't think about it. I am promoting national interests, and that is the narrative I go with, that is what I use to promote its [research's] value"

Researcher (Professor)

- Some grimpact is foreseeable, but researchers choose not to acknowledge it because it is not part of their narrative.

- Motivations for this choice is competitive compliance, rather than a commitment to a truth

Linearity in Impact pathways

Science space

- Ethics
- Research integrity
- Open Science
- Co-production
- Stakeholder involvement

"Ooops allowed"
Experiments encouraged
Learning through errors. Can amend
Control of dissemination

Societal space

"Ooops not allowed"
Uncontrolled dissemination
Uncontrolled experimentation
Experimentation/errors/learning = weakness
Unanticipated actors/stakeholders

- Stakeholder use
- Serendipity
- Implementation
- Freedom

Assessable/
Rewardable/ Valuable

Research is rewarded for scale, speed and visibility but not for downstream harm, fragility or misuse

AI doesn't create societal risk, its fast adoption exposed how we reward ignorance of the risk

Consequential/
Inconvenient


Grimpact is the space between research impact and research reward, where the nature and value of societal benefits are in flux and continually contested

Academia rewards

- Novelty
- Performance gain
- Adoption and uptake
- Police relevance
- Benefits
- Change

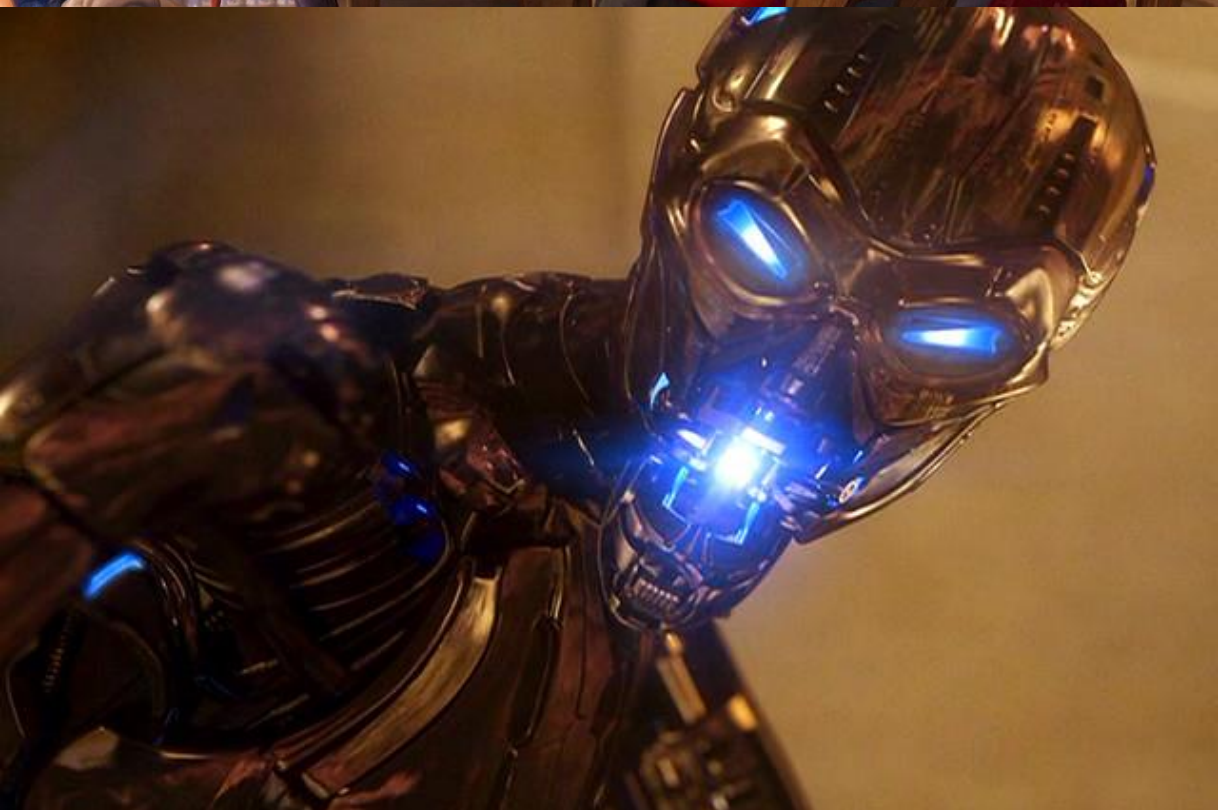
Grimpact by design

Academia ignores

- Who bears the risk
- Who loses trust
- Who absorbs environmental and social costs
- Who is accountable when things go wrong

When failures happen, trust collapses – fast!

People blame institutions NOT algorithms
E.g. Automated welfare and immigration decisions; predictive policing and risk scoring; AI-mediated public services



The Terminator warned us about losing control

The Mitchells warned us about giving control away willingly.

AI's societal risk persists not because we don't see them, or haven't been warned
BUT because our public knowledge structures don't reward researchers for slowing down, saying no or naming harm.

The question is no longer whether AI poses societal risk but whether our research systems are willing to recognise
their role in producing it.


Thank
you!

Characteristics of negative impact (initial study)

1. **Violation of 'normal' impact**

Grimpact is characterised by the absence of normal impact, and the distinction between the researchers and the subjects of research.

2. **Attribution (who is to blame)**

Attribution is more difficult in Grimpact than it is from impact.

"blame" and "fault" versus "duty" and "success"

3. **Grimpact is contagious**

It acts fast, invades other fields and beyond the geographical scope of normal impact research-user relations

4. **Misconduct not always necessary – insofar as restitution in science space not sufficient**

Research misconduct OR a transgression between acceptable academic and non-academic behaviours including strict ethical controls that govern researcher behaviour.

Grimpact in other forms

- Post truth (MacIntyre, 2019)
  - *How we arrived in a post-truth era, when "alternative facts" replace actual facts, and feelings have more weight than evidence.*
- Uncertainty in policymaking and unintended consequences (Oliver et al, 2019)
- Epistemic risk (Sahlin & Persson, 2014)
  - Uncertainty due to gaps in knowledge, and the risk of different forms of knowledge being used to provide explanations
- Implementation gap risks
  - Morally targeted use of evidence (Haynes & Derrick et al, 2011) increases the risk of Grimpact
- Normal versus extraordinary research
  - Normal impact *is found in the responsible relations between academia and other institutions of civilization* (Sivertsen, 2018)
  - Daily activities and how they are organized
  - Normal impact is directly involved in markers of civilization such as higher education, and freedom of press
  - Normal impact can change to having extraordinary impact.
  - Example, Syrian-Norwegian collaboration on Palmyra (UNESCO word heritage site) since 2008
- Responsible research and innovation
  - All interventions and assurance concentrated downstream in the 'science space'