AI EXTENDERS:

THE ETHICAL AND SOCIETAL IMPLICATIONS OF HUMANS COGNITIVELY EXTENDED BY AI

José Hernández-Orallo
Universitat Politècnica de València, Spain
Leverhulme Centre for the Future of Intelligence, UK
jorallo@dsic.upv.es

Karina Vold
Leverhulme Centre for the Future of Intelligence
University of Cambridge, UK
kvv22@cam.ac.uk

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

CFI LEVERHULME CENTRE FOR THE FUTURE OF INTELLIGENCE

AAAI / ACM Conference on Artificial Intelligence, Ethics and Society, Honolulu, 27-28 January 2019

WHAT IF AI WERE PART OF YOU?

- How would the ethical issues be considered, e.g., ***you*** becoming unfair?
- What risks would ***you*** create (to others and yourself)?
- How would ***you*** impact society?
- Who would be responsible for ***your*** actions?
- How to see a “human in the loop” if AI were in the loop for ***your*** decisions?
- What are the rights about the access, use and disposal of a part of ***you***?

AI is already becoming part of you

THE EXTENDED MIND HYPOTHESIS

- “Tools” can serve as partially constitutive ‘extensions’ of an agent’s (e.g., a human’s) cognitive states (Clark and Chalmers 1998).
- “Parity argument”: we should treat computationally equivalent processes with “the parity they deserve”, irrespective of whether they are internal or external to the skull.
  - Traditional example: pen & paper. Nowadays: an interactive drawing gadget

What if these processes are accomplished by a cognitive tool using AI?

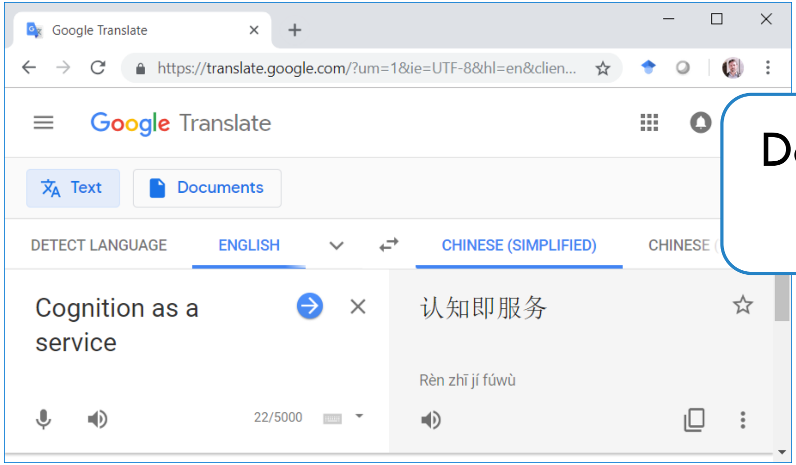
HOW CAN AI BE USED AS A TOOL, FOR AUGMENTATION?

- Autonomous AI:
  - Most common interpretation: AI as an agent!
    - They do perform tasks on their own

Dominant narrative: the whole process is automated (replaced)

HOW CAN AI BE USED AS A TOOL, FOR AUGMENTATION?

- Non-autonomous AI:
  - Externalized cognition (A ← E): an outsourced service
    - Cognition as a service (Spohrer and Banavar 2015), AI services (Drexler 2019).



Cognition as a service

认知即服务

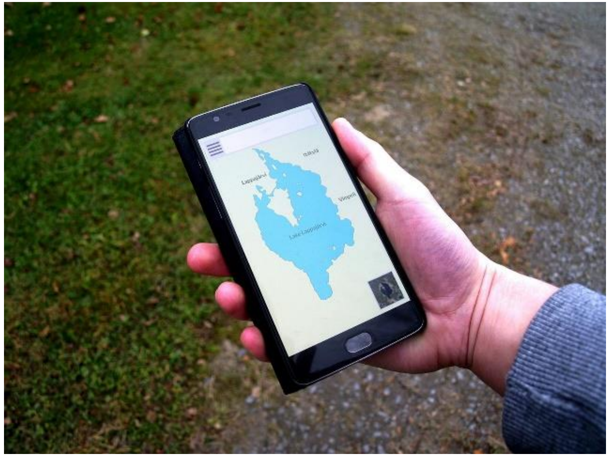
Rèn zhī jí fúwù

22/5000

Dominant narrative: partial automation (subprocesses are replaced)

HOW CAN AI BE USED AS A TOOL, FOR AUGMENTATION?

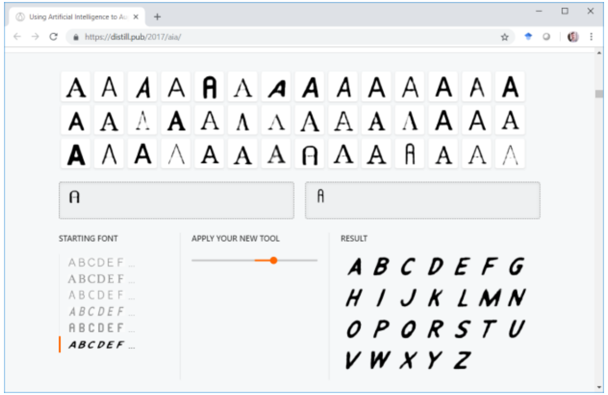
- Non-autonomous AI:
  - Extended cognition (A[E]): highly coupled
    - The tool is always needed, the process is not internalized (at most the interface)



Dominant narrative: people are empowered

HOW CAN AI BE USED AS A TOOL, FOR AUGMENTATION?

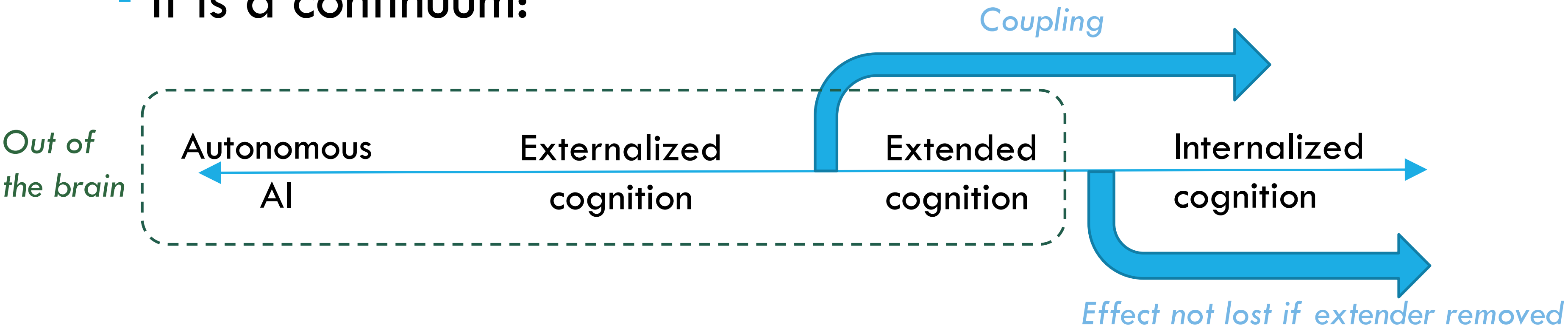
- Non-autonomous AI:
  - Internalised cognition (A^E): done externally, but then mimetized internally
    - "computers are a means to change and expand human thought". Carter & Nielssen (2017)
  - AI generating culture (new words, concepts, ideas, representations, etc.)



Dominant narrative:
people are enlightened

A RANGE FROM AUTONOMOUS TO INTERNALIZED

- It is a continuum:



Many processes can't be internalized (resources, AI interpretability, ...) but they can be extended or externalized.

AI EXTENDER

- We refine Hutchin’s (1999) definition as follows:

> A cognitive extender is an **external** physical or virtual element that is **coupled** to enable, aid, enhance, or improve cognition, such that all – or more than – its **positive effect is lost when the element is not present**.

  - **AI extenders** are cognitive extenders where some of the cognitive processes require the extender to have AI.
  - Extenders are more **human-centered** than human-like. They may personalize the best cognitive aid depending on the extendee and situation.

WAYS AI CAN EXTEND COGNITION

- Memory processes
- Sensorimotor interaction
- Visual processing
- Auditory processing
- Attention and search
- Planning, decision-making and acting
- Comprehension and expression
- Communication
- Emotion and self-control
- Navigation
- Conceptualization
- Quantitative and logical reasoning
- Mind modeling and social interaction
- Metacognition

IMPLICATIONS: GOOD OR BAD IN A DIFFERENT WAY

- Analyzed for augmentation:

⊕ Positive effects narrative: very empowering

(Bostrom and Sandberg 2009)

⊖ Negative effects narrative: potential risks

(Carr 2015, Frischmann and Selinger 2018, Danaher 2018, Carter and Palermos 2019)

But these effects not analyzed under the extended cognition thesis

(e.g., Danaher, 2018, views social interactions using augmentation as potentially deceptive).

IMPLICATIONS: ATROPHY AND SAFETY

Navigation system fails in the middle of nowhere!

- The positive effect is (more than) lost when the element is not present
  - It creates dependency
  - Cognitive miserliness (Barr et al. 2015) leading to “degeneration” (Carr 2015)
  - Atrophy also generates many safety issues

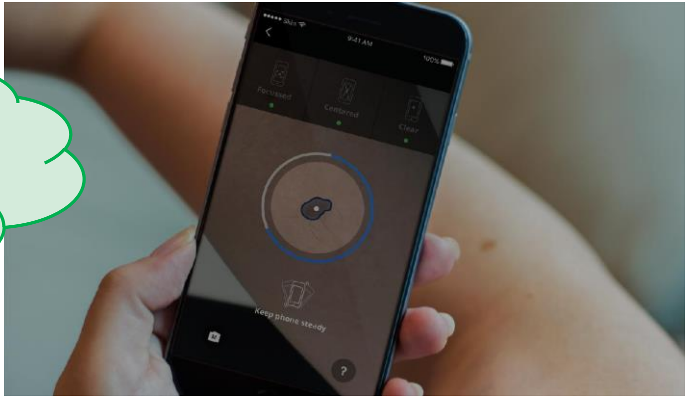
IMPLICATIONS: MORAL STATUS AND PERSONAL IDENTITY

A artist and their creative gadget

- Heerminsk (2017): “degree of dependency and integration [is] proportional to the artifact’s moral status”.
- What A[E] creates should be owned by A, not E (Frischmann and Selinger 2018). Actually, it should be owned by A[E].

IMPLICATIONS: RESPONSIBILITY AND TRUST

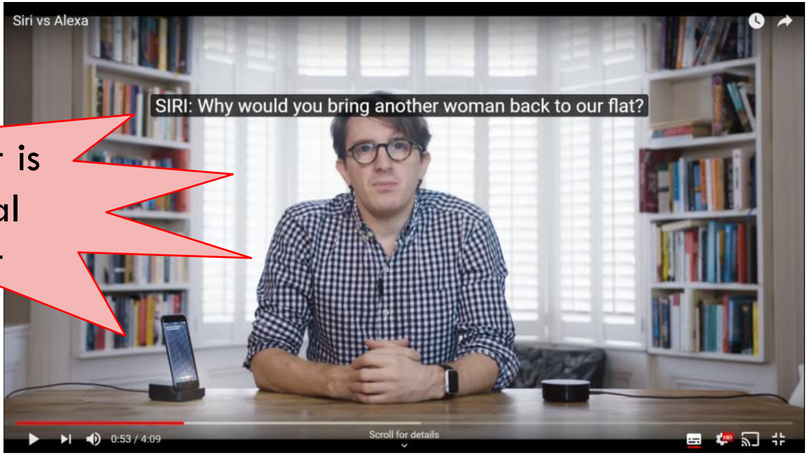
A doctor and a malignant mole detection device



- Disclosure clauses inherited from software totally unacceptable!
- Developers of an extender E can use the extendee A as a puppet:
  - Tempting when E is not allowed to operate on its own (human-in-the-loop).
  - E dominates in A[E], but leaving the responsibility to A.
- Should patients trust the doctor or the device?

IMPLICATIONS: INTERFERENCE AND CONTROL

A personal assistant is "jealous" of its rival personal assistant


Siri vs Alexa

SIRI: Why would you bring another woman back to our flat?

- The extender E monitors the human A and looks for interventions that empower A.
- This may degenerate into sophisticated ways of surveillance and manipulation, well beyond nudging.
  - Not only do they modify beliefs, but ways of thinking!!

IMPLICATIONS: EDUCATION AND ASSESSMENT

Should we hire/admit A1[E] with score 7.3 or A2 with score 7.1 ?

PROFESSIONAL REGULATION COMMISSION
PROFESSIONAL REGULATION COMMISSION • REPUBLIC OF THE PHILIPPINES •
UPDATED LIST OF ALLOWED CALCULATORS
depedlps.blogspot.com

- Is it fair (and meaningful) to remove the extenders when assessing how fit a person is for a job?
- Would the extender be acceptable in the real situation?
- How can we evaluate accurately with the extenders (e.g., avoiding cheating)?
- How to be sure that the extenders will be the same in operation?

RECOMMENDATIONS

- The nature and impacts of AI extenders are sufficiently distinctive to treat them separately from other extenders and other kinds of AI.
  - **AI developers** should take care to distinguish when they are developing an autonomous system, a decoupled system or a fully-coupled system.
  - **Philosophers of mind and psychologists** viewing AI extending human cognition may reconsider what is potentially dangerous or unethical.
  - **Regulators and policy-makers** should not only regulate autonomous systems, but extenders as well: humans may be "used" to circumvent these regulations.