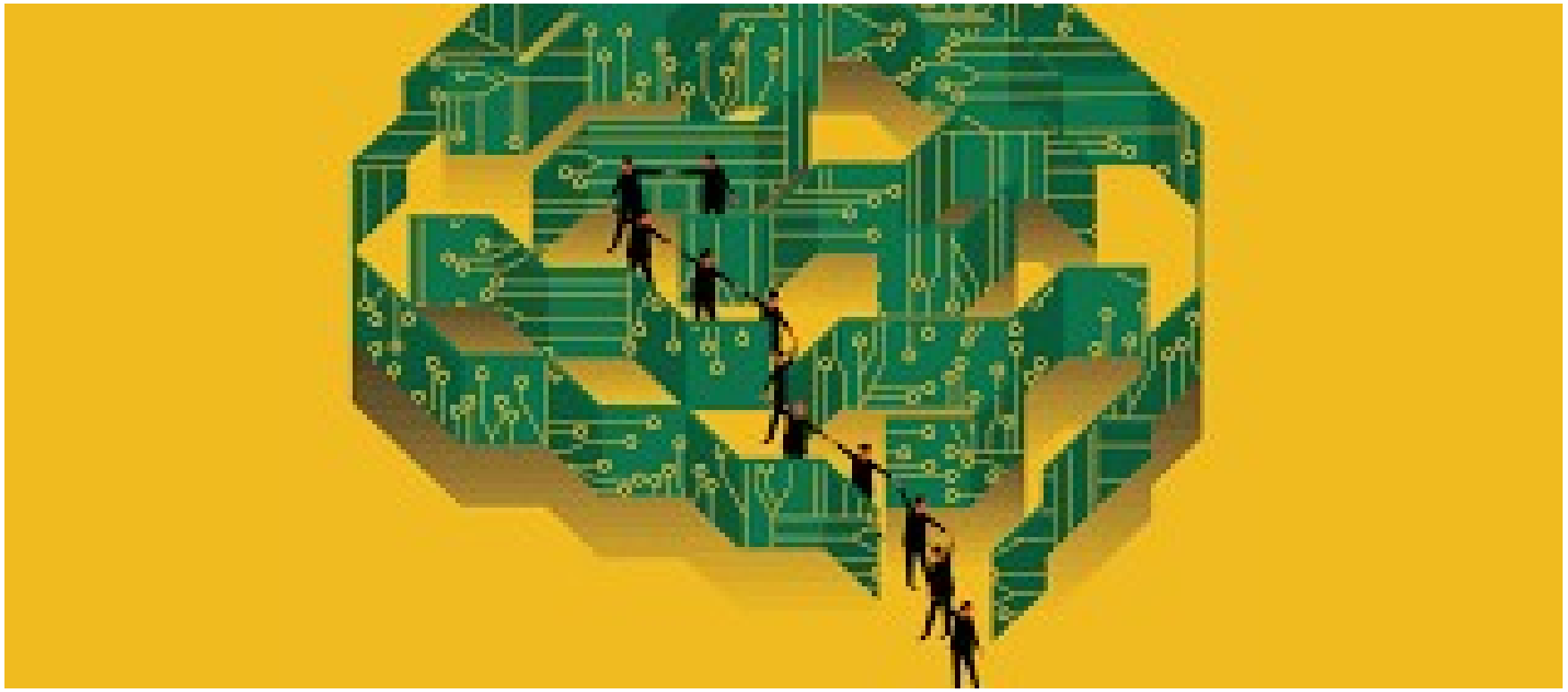


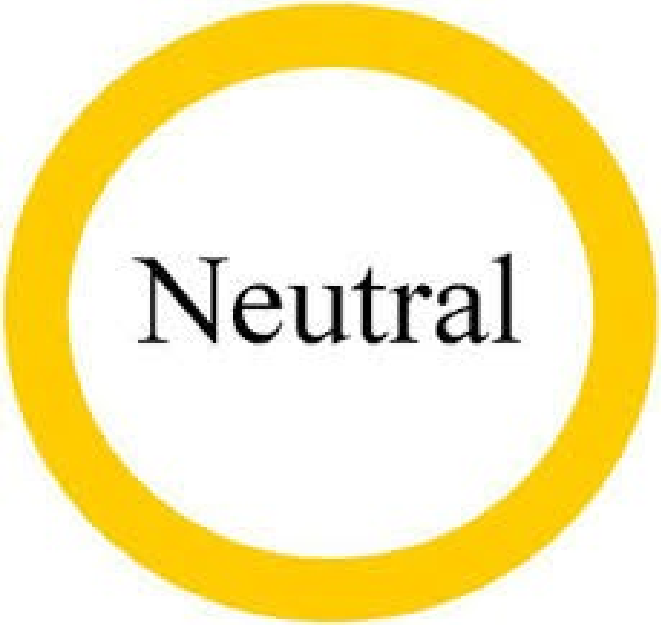
Embedded values and value sensitive design

Embedded values

- Claim: all technologies implicitly embed values (and biases)
  - E.g. goals, context of use, assumed user characteristics, specific technical implementation
- “the design and operation of computer system has moral consequences and therefore should be subjected to ethical analysis” (Brey 2009)
- it is possible to identify tendencies regarding promotion of particular values and norms
  - for example, computer programs can be supportive of privacy, freedom of information, property rights or go against the realization of these values.

Technologies are not neutral

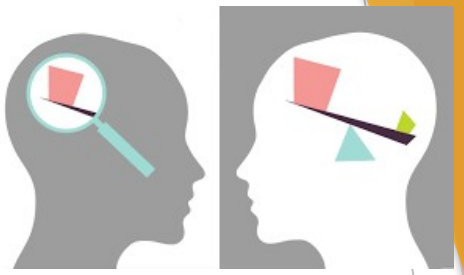
- Counterposition to those who focus on values only post-technology development, i.e. consider technology design to be ethically neutral, and only its application to be subject to ethical considerations
  - E.g. common view of engineering as neutral service-provider for client who is responsible for ethical use
- Importance because: information technologies often not transparent to users, i.e. users cannot easily investigate technologies and their biases



Questions with regard to embedded values

- What specific **values** does a technology embed?
- Who has **control** with regard to how the technology functions in relation to value concerns? What role is given to the **user**?
- Is a technology **transparent** or opaque?
- Does the technology support **balanced** terms of information exchange between **technology provider and technology user**, or is one side only benefitting at the cost of the other?
- Does the technology **unfairly discriminate** against specific sectors of potential users? (e.g. socio-economic, disability, gender?)
- Does the technology enhance or diminish **trust**?

Biases in computer systems



- Definition: A system that systematically and unfairly discriminates against certain individuals or groups in favour of others
- Denying opportunities or assigns undesirable/negative outcomes:
  - E.g. credit rating system that instead of focusing on actual credit record bases its assessment on address or ethnic surname
  - Predictive policing targeting minority populations that will disproportionately be affected by police scrutiny (e.g. minor drug possession charges when searches)
  - Criminal justice parole assessment systems that discriminate against minorities
- Particularly problematic if no possibility of appeal/correction or validation and if a system becomes a standard (monopolist) in the field and no alternatives exist

Value sensitive design (Friedman, Kahn & Borning)

- Value-sensitive design refers to a theoretical approach that identifies and incorporates stakeholders' values into the design process of a new technology
- It is 'a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process' (Friedman, Kahn & Borning 2008).

Procedure of value sensitive design

- A combination of 3 types of investigations as part of the design process:
  - **Conceptual**: identifying relevant values and who might be relevant stakeholders
  - **Empirical**: actively engaging stakeholders' perspectives to identify variation and commonalities
  - **Technical**: proactive design to support values identified in the conceptual investigation; identify how existing technological properties support or hinder realisation of values

Common characteristics of ethics by design (SIENNA 2020, Donia & Shaw 2021)

- Identification of a one or several core principles or value categories relevant to technology domain
- Application to the specific technology, specifying ethically relevant features (e.g. as “ethical requisites”)
- Consultation with relevant stakeholders
- Practical design decisions to realise ethically relevant features
- Identification of relevant parameters of the application context, to support ethical design as well as deployment and use

Who is responsible?

- The impact of information technology on society depends on various stakeholders: e.g. technology developers, the organisation/company that tasked them, users, regulators etc.
- Design ethics approaches seem to focus on responsibilities of developers – is that where responsibility should lie?
- Attempts to provide ethical training or reflection opportunities tailored to developers as one option (e.g. harnessing existing software development processes, Zuber et al. 2021, or using “Moral IT card deck”, Urquhart & Craigon 2021)
- E.g. Donia & Shaw (2021) argue that a realistic understanding is needed of contextual impact and limitations on developers’ agency in development process

When to apply comprehensive ethics by design approaches?

- Proportional approach
- Especially when new technologies are developed that open up **new ways of impacting human life**
- **Proactive** approach rather than reactive approach, i.e. identify issues before they occur rather than having to troubleshoot later
- Goal is to use technology to **meet complex human needs** and open up new possibilities of **"living a good life"** rather than restricting them or forcing practices on unwilling adopters

Example: Care robots

Privacy by design (Cavoukian)

- A method of embedding privacy concerns into the design of technology
- Privacy as prominent concern in information technologies (“privacy policies”); respect for privacy as crucial value – and legal obligation (GDPR)
- Based on recognition that privacy interests must be addressed proactively
- Early consideration of privacy in design process, to facilitate system functionality with privacy, rather than imposing privacy constraints on functioning system

7 Principles of Privacy by Design

- Proactive not reactive, preventative not remedial
- Privacy as default setting
- Privacy embedded into design
- Full functionality - Positive sum, not Zero sum
- End-to-end security – Full lifecycle protection
- Visibility and transparency – Keep it open
- Respect for user privacy – Keep it user-centric