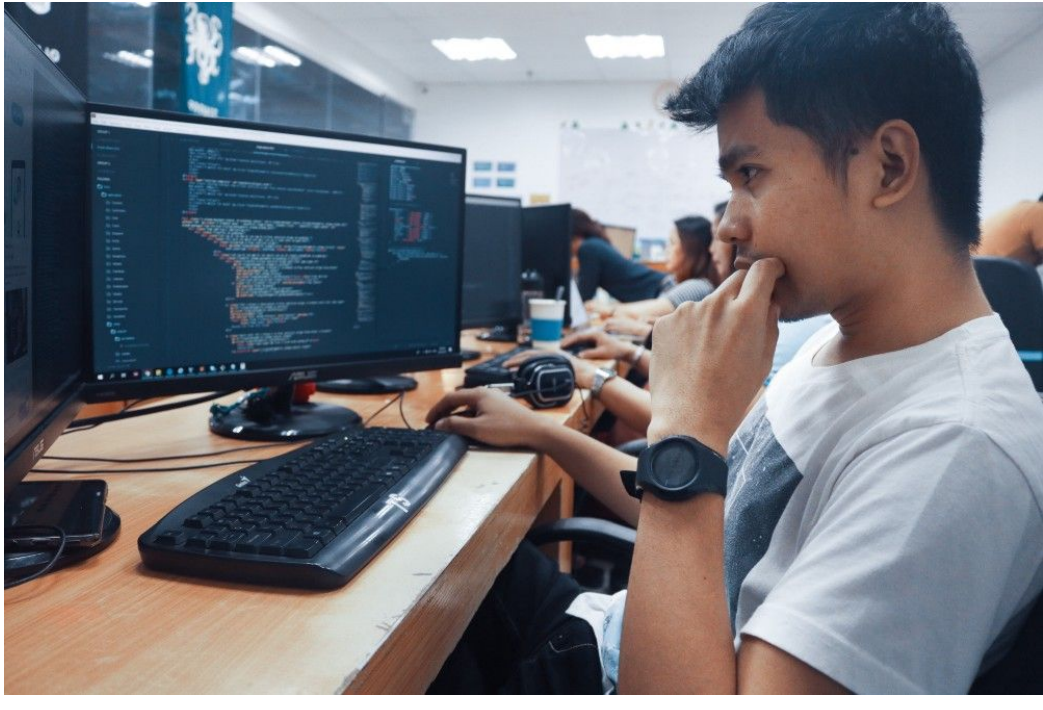
# C6 Moral AI
# C7 Philosophical AI

Daniel, Ludovic, Pasquale, Souad, Yixiao

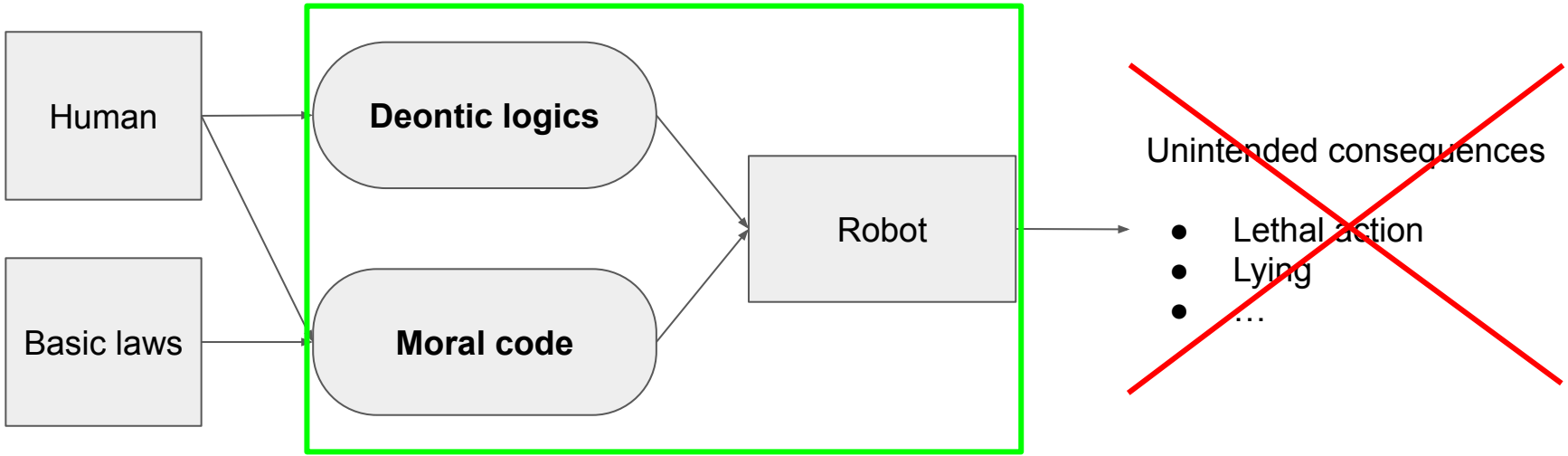
# 6. Moral AI – Definition

Aka robot/machine ethics, ethical AI, moral robots → autonomous decisions

Computer ethics → human decision



# 6. Moral AI – Framework




**Moral code**: required for ethical reasoning

**Deontic logics**: Follow moral code under all circumstances

# 6. Moral AI – Future

Machine ethics in infancy

Will grow with smarter AI

Human ethics ⇏ machine ethics (we hold machines to a higher standard than humans)

# 7. Philosophical AI

≠ philosophy of AI

Daniel Dennet's (1979) views:

**Claim**

- AI *is* philosophy
- AI attempts to explain intelligence by designing and implementing abstract algorithms that capture cognition (top-down approach)

**Rebuttal** "AI is AI, not philosophy;  but it's AI rooted in, and flowing from, philosophy"

# 7. Philosophical AI – Fatal flaws

1. AI is attempt to substantiate thesis that intelligence is computational (machine Turing-level)
   Philosophical claim ⇏ philosophy

2. **Claim**: \(f\) computable ⇔ \(f\) Turing-computable
   **Flaw**: information processing can exceed Turing-computation (*hypercomputation*)
   AI constrained by "mentation consists in computation" (*mechanistic* solutions).
   Philosophy and psychology are not.

3. AI researchers and developers are not philosophers!

# Moral AI

•Moral AI (also known as ethical AI, machine ethics, moral robots, or robot ethics) applies in situations where robots are able to take autonomous decisions.

•It should not be confused with Computer Ethics which contemplates situations where the decision maker is a human being.

•Moral AI is needed not only in machines that can perform lethal actions but more generally applies to a wide range of situations and machines (so called moral machines).

# Approach to Moral Machines

•Machines that can reason ethically via the implementation of a moral code.

•The code can be sourced by humans directly or inferred by the machine itself.

•The machine will have to follow the code under all circumstances and the following of the code must not produce an unexpected outcome.

# Deontic logics

- The approach used to make sure the application of the moral code will not lead to unwanted consequences is called deontic logics.
- Deontic logics are clear under any circumstances, meaning for any problem there is always only one clear solution

# Other references and citations

Lyan Watson:

“If the brain were so simple we could understand it, we would be so simple we couldn’t.”