Topic:
Approaches and Architectures: Bottom-up vs. Top-down

Structure:

- What is Machine Ethics? (4, 5)
- What categories of approaches are there? (1, 4, 5)
- Comparison: Advantages and Disadvantages (1, 4, 5)
- Examples:
  - Top-down: LOGIKEY (2)
  - Bottom-up: CIRL (3)
- Discussion (based on the question: What category is best?)

References:

1) Wallach, Wendell & Allen, Colin & Smit, Iva. (2008). Machine Morality: Bottom-up and Top-down Approaches for Modeling Human Moral Faculties.
2) Benzmüller, Christoph & Parent, Xavier & van der Torre, Leon. (2019). Designing Normative Theories of Ethical Reasoning: Formal Framework, Methodology, and Tool Support.
3) Dylan Hadfield-Menell, Anca Dragan, Pieter Abbeel, and Stuart Russell. (2016). Cooperative Inverse Reinforcement Learning.
4) A. F. Winfield, K. Michael, J. Pitt and V. Evers. (2019). Machine Ethics: The Design and Governance of Ethical AI and Autonomous Systems [Scanning the Issue].
5) https://www.ethicsnet.org/blog/Blog/approaches-to-ai-values