Oral Question

Lesson 1 (Castelfranchi)

- main problems in mass medias

Lesson 1 (Sartor)

- 4 ethical principles of the guideline for trustworthy AI
- Human rights in chapters in ethical AI (inside Trustworthy AI:
  fairness (being equally treated, freedom expression, dignity, privacy and data protection, data of property of data, transparency, explainability)

Lesson 2 (Utilitarianism & Morality)

- Consequentialism
- What is utilitarianism
- Consequentialism and deontology in ethics, and utilitarianism
- Utilitarianism in general, is a good guide for AI ethics (take in consideration utility of everybody)

Lesson 3 Artifacts' Politics & Resp and Automation

- What do we mean with technological mediation?
- With mediation we can introduce some biases in the technology?

Lesson 4 (Deontology)

- Kantian ethics (golden rule, hypothetical and categorical imperative)
- Kantian approach in ethics

Lesson 5 - Ethical Knob & Value Alignment

- Ethical knob (AV avoid collision, knob to set relative importance between passenger and pedestrian, knob set by passenger for freedom and manifactur can't done global choice for all

- Value alignment (enforce AI to follow ethical rule, such as ethical knob)
- Reward hacking and how can we link chatbot Tay with this phenomenon?
- reward hacking
- Bias and how to counteract it
- Flexibility of cp-nets vs bottom-up approaches

Lesson 6 - Human Rights & Logic Programming

- Definition of human rights
- Explainability, legal references (symbolic AI more explainable than subsymbolic AI, a model which give logs and explaination is better)
- Is logic the only way to implement utilitarianism?
- How to implement utilitarianism principle: logic programming, fitness function?

Lesson 7 - Modelling Norms

Lesson 8 - Data Protection

- GDPR
- Right of explanation into GDPR: users have to know why content filtered out
- Profiling
- Data protection by design (and where this principle is mentioned)

Lesson 9 - Fairness in Algorithimic Decision Making

- Compass does not take into account race as feature, but black people were poorer etc so were more criminal, unbalanced outcome beacause of it
- Compass (static and dynamic, several features, it is fair if it not consider race, religion), is compass fair for you

Lesson 10 - Autonomous Vehicles

- Describe the ladder of autonomy for autonomous vehicles
- Technical and ethical problems of AV

Lesson 11 - Claudette System

Lesson 12 - Intelligent Weapons

- Principles of ius in bello
- What is liablity gap in AWS (now impossible to give responsibility to a machine)
- Human control in AWS (human in-out the loop)
- Jus in bello and legal issues for AWS (discrimination, proportionality, necessity), AWS now does not guarantee these
- Principle of distinction in targeting, ius in bello, principle of humanitarian law, principle of necessity (kill acceptable if you see an enemy, problem in AWS is to identify who is combattent who is civilian)
- Are there domain where AWS should be legal to use? (yes, some protect military basis from attacks, so defend)
- Human dignity involved in AWS (not respect in target detection enemy vs civilian)
- Main principles of laws of war, responsibility gap
- Human rights in debate of autonomous weapon system, basic rules of rule of war (principles)

Lesson 13 - Ethics of Filtering

- Issue of filtering
- Eco chamber and filter bubble
- Principle of ethics of filtering such as freedom of expression (some sexual and agressive content is ok to filter out)
- Automatic tools for filtering? (Yes, no moderator, ...)
- Moderation

- Human rights in ethics of filtering (freedom expression, cultural expression, respect of minority, data privacy, reputation)
- Human right issues in filtering (freedom)
- AI can contribute to filtering
- Main legal problems of filtering