Mid-Term Project: Annotated Lesson Plan Template

Student Names: Luka Blagojevic, Liam Ryan, László Kőszeghy

Course Name: Interdisciplinary Perspectives on Artificial Intelligence

Topic of the Lesson: AI and Consciousness: Can a machine be a mind?

Part 1. Reflection Questions

1. ***Teaching philosophy*** *- What is your approach or philosophy to teaching in short (what do you **value**) and how is this demonstrated in your lesson?*

   We generally take a constructivist approach in this course, however we note that given the interdisciplinary nature of this course, some class may adopt different approaches. Our constructivist approach entails that while we have specified pre-set learning outcomes (see below), these are achievable *only through* arriving at meanings which emerge via discussions during class. In other words, we believe that the back-and-forth exchange of ideas is a *constitutive* element of learning: we couldn't "convey" or "transfer" a pre-existing set of propositional knowledge that would capture our envisioned learning outcomes (again, there might be exceptions). We want our students to learn to engage in philosophical discussions around one of the most pressing contemporary issues: that of the development of AI.

   Therefore, our syllabus contains many shorter readings/viewings whose function is to prompt students to think independently. These exciting prompts (movie/video excerpts) at the beginning of each session will help us get off the ground and allow students to feel comfortable contributing in a general way. In other words, an important principle that guides our building our syllabus is that we do not necessarily want our students to read all the important readings there are about AI – instead, we want to them do the intellectual work of thinking through, on their own, some of the main ethical/political/ontological/mechanical problems related to the topic.

   We start out with the assumption that our students are, with the right amount of scaffolding and support, capable of formulating their own viewpoint, given this is an elective later-year undergraduate course. Thus, we value independent thinking. The classroom, accordingly, will be a space for active engagement where students can feel safe to contribute to fundamental philosophical problems. We will encourage our students to contribute and thus experience the joy of thinking together as a community. The theoretical nature of the session and the whole course is well suited for active discussions.

   However, we will not merely focus on theory. We believe that students should be taught in a way that they can connect the classroom knowledge to the real-world world phenomena. This is demonstrated by providing real-life contemporary examples, even for the most theoretical problems in philosophy of mind.

Furthermore, the reasoning and critical speaking, and writing skills acquired during this session and throughout the whole module are crucial transferable skills to other disciplines.

2. ***Context.*** *Consider the name of the course, degree level (e.g., MA), and students (number and background), and how this course fits into the student's degree program (e.g., mandatory methods course or interdisciplinary elective).*

   Course name: Interdisciplinary Perspectives on Artificial Intelligence

   This is an elective course targeted for undergraduate students in (open to interested MA students), cross-listed by the Dep. of Philosophy and the Dep. of Computer Science.

   The expected number of students is 15-20. No special prerequisites or background is required, although completion of the first year of studies is encouraged. We expect a mixture of students from both departments and even some from other departments such as mathematics.

3. ***Scaffolding.*** *What is the context of the lesson in the course, and how will this link to previous (and upcoming) classes in the semester?*

   This lecture is within the multidisciplinary course offered by the Dep. of Philosophy and the Dep. of Computer Science: Interdisciplinary Perspectives on Artificial Intelligence.

   Previous lectures include:

   - Introduction to AI: Machine Learning Concepts
   - Neural networks, AI & Deep learning

The lecture would not heavily rely on these previous sessions, as it is situated within an interdisciplinary course, which should mostly provide breadth of topics. The aim is to provide the students with the variety of advancements caused by the developments in the field of artificial intelligence and the issues that arise with it. This means that each lesson should be considered as a self-contained, where all the relevant background knowledge is introduced in the lesson itself.

The " AI and consciousness: Can a machine be a mind?" session should provide conceptual and philosophical tie for the future interdisciplinary lectures on AI, such as:

- AI in Social Media Landscape
- Ethics of AI
- Military Application of AI

- Virtual Reality and AI

4. ***Learning outcomes.*** *What do you <u>really</u> want students to learn? Based on the above, please formulate 2-3 <u>SMART Learning Outcomes:</u>*

We will be employing the **SMART** methodology for measuring outcomes:

Specific

Measurable

Attainable

Realistic

Trackable

Thus, by the end of this session, students:

- Will be equipped with some of the basic concepts that can be employed to tackle arguments about whether or not computers can be minds. They will also be equipped with skills to employ interdisciplinary techniques to their respective departmental learning;
- Will learn about one influential argument concerning the difference between AI and the human mind, the Chinese Room Argument, and understand the arguments and implications. Students will be able to articulate one theory of mind (functionalism);
- Will understand the current state of AI development with regards to its functionality and its potential developments into the future.

Evidence for this learning will include, but not be limited to:

- Formulate an overview of, and argument for or against, functionalist theories of mind.
- Formulate an argument for, or debate whether, AI is a realistic possibility, and understand what the argument entails ethically/socially.
- Formulate an argument for/articulate one theory of mind (functionalism) as it relates to the plausibility of AI and understand what the argument entails ethically/socially.
- Formulate an argument for/articulate one contemporary technology that is relevant to the development of AI and understand what the argument entails ethically/socially.
- Give real-world examples of AI that have specific functionalities and connect articulate theoretical concepts through these examples

- Articulate open questions moving forward for the future of AI, mind, and ethical applications.

5. ***Preparation.*** *(Students) What do students need to **do** to prepare (e.g., reading)? (Students): What threshold concepts or knowledge do they need to participate fully in the session? (Instructor) What do you need to prepare as the lecturer (e.g., polls, reading)?*

   We as instructors need to carefully prepare video material for the Part 1 and the Part 3 of the lecture. This means video editing, downloading material, and inserting it into the presentation. We need to send out preparatory material as well. We will also need to be familiar with the material.

   Timeline in relation to full course

   2 Weeks before term starts: latest version of syllabus uploaded to the Learning Management site, along with readings/required materials
   1 Week before the session: reminder email about required materials (readings, viewings)
   3 days before the session: videos to be shown in class edited, embedded in PowerPoint presentation; PowerPoint also ready
   20 minutes before the session: PowerPoint (with sounds and video) tested and projector set up in the relevant room/Zoom

   For the session, students will have to read/watch the following:

- Boston Dynamics. (2021, August 17). *Atlas | Partners in Parkour*. https://www.youtube.com/watch?v=tF4DML7FIWk
- *Can we make consciousness into an engineering problem? | Aeon Essays*. (n.d.). Aeon. Retrieved November 30, 2021, from https://aeon.co/essays/can-we-make-consciousness-into-an-engineering-problem
- *Minds, brains, and programs | Behavioral and Brain Sciences | Cambridge Core*. (n.d.). Retrieved November 30, 2021, from https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/minds-brains-and-programs/DC644B47A4299C637C89772FACC2706A

   In order to ensure that students do the required reading/viewing and to foster critical engagement with the material, before class, each student will have to submit their **"reader's diary"** with critical reflections and select one question which they post it on the given learning management system (I.e. Moodle). The description of the task is as follows:

- This assignment is to help you understand the assigned readings. The exact format of your reader's diary is up to you, but **your notes need to provide evidence of your having read and understood the required readings**. It is highly recommended to use

the **split page note-taking method**: on one side of the page, you should give a **list of the argument(s), main point(s) new terms, or key concepts that you find in the reading** (you can use bullet points and a note format), and, on the other, **provide your own critical reflections.**

- Effective note-taking will help you develop the argument that you will want to make when writing your essay(s). Also, the right quote that you will want to use weeks from now, could easily be lost if you do not record it. The reader's diary is there to help you think and keep track of the process through which your thinking develops.

  After submitting their reader's diary, students choose and post one critical/clarificatory/provocative question/short reflection related to the reading in the forum of the learning management platform.

  After class, for homework, we will ask them to provide an answer to their question or how has their reflective opinion has changed, drawing upon what they will have learned that day (they can also choose one of the post-class poll questions [see below]). E.g. "I didn't understand the `other minds reply' in Searle's *Minds Brains and Programs.* Why exactly is this a problem for AI?" And the response "There is a difference between what we can attribute to other minds, such as characteristics of mind, and whether we can know for certain that something has a mind. These are not the same thing."

6. **Anticipating Challenges.** What are the anticipated bottlenecks or challenges, and how will you address them?

   Part 1 is fun with easy-to-answer and non-threatening questions (date/romance with robots). Some students might be shy or reluctant to share their subjective opinions on this topic (since it covers romantic behavior). In that case, questions will be rephrased in a manner that emphasizes communications aspects of the conversations - "Okay, regardless of the Will Smith and their date, how does the robot appear to you? Does it sound human?". If there is still some inactivity, the teacher should start giving his own examples, one by one, to slowly encourage students to participate with their own examples.

   Part 2 dives into more theoretical waters. We will have to do the necessary scaffolding and ask easy-to-follow guiding questions to help students with the bottom-up thinking process of understanding what general principles follow from our discussion in Part 1. Searle's paper can be challenging for some students, so we need to take care not to rush the discussion. In case there is a lack of activity or the questions prove to be too hard for the students, the lecturer will start over and explain each concept carefully while trying to assess if the students are following him. After each concept he may ask a question to check if the students have understood it and try to encourage them by tackling smaller problems first. In the worst-case scenario (where students cannot follow the ideas), the professor will slowly explain the concepts, pose rhetorical questions in place of students ("You might be wondering, why is..."), answer them until the end of the Part 2.

When listing answers on the whiteboard in Part 1 (discussing why artificial intelligence in the clip can/cannot be considered a human), students might not give answers that we have expected, thus making it harder to transition later into Part 3 (where we connect these concepts to various video examples). If this scenario occurs, we will add the missing ideas at the end with some basic explanations of why we added them. Some answers to the question - "What do humans have or they can do, but AI cannot". which are important to write down on a whiteboard:

- Embodiment/Movement

- Communication/Language

- Problem solving

- Teamwork

- Creating Art

- Normative thinking

These ideas are directly tied to the videos that we are going to show (they are in the materials in the table below).

The IT equipment might not work properly, which would be a negative impact, since there is a lot of video examples from the start of the lecture. To save the lecture, the material will be sent out at that time to everyone, so people can watch in on their phones or laptops in small groups.

7. ***Assessment.*** *How and when will you assess whether students learned or acquired the skills what you intended?*

During the course session, we will constantly engage students with questions and thereby assess whether they can follow the discussion. At the end of each session there is a post-class poll with an additional question such as "What was most intriguing for you during this session?", which is, on the one hand, fun for students (see post-class poll questions below), on the other, gives us feedback about what students have grasped of the material covered during the class.

For the upcoming session, they will have to write a short, 300-400-word-long piece answering either their own pre-class reader's diary question, *or* one of the prepared post-class poll questions.

At the end of the whole course, students will have to write an 1500-1800-world long paper. They will be able to choose one of the four essay prompts provided. The four essay prompts will be quotations from philosophical works reflecting on the main

topics covered during the course. The essay prompt pertaining to this session is the following:

"The reason that no computer program can ever be a mind is simply that a computer program is only syntactical, and minds are more than syntactical. Minds are semantical, in the sense that they have more than a formal structure, they have a content."

John Searle

The prompts will be closely related to the readings students analyzed in their reader's diaries and post-class poll questions. Thus, they will already have a connection to the essay prompts and will have prepared in advance for the assessment.

Part 2. Lesson Plan Activities

Using the table below, please create a detailed description of the activities of the lesson, including timing, exact questions, precise instructions, materials to be used, and major points to make in any lectures or discussions. Please feel free to add additional rows if needed.

Time	**Activities** (including format, objectives, exact questions, precise instructions, and major points to make in any lectures or discussions)	**Materials & Preparation**
20	**Part 1 – Introduction** **Overall guiding question for class:** - Start lightly with some ideas and definitions from the previous sessions, as a sort of quick recap (what did we do last time? What's an AI from a technical standpoint, how do these algorithms work on a basic level) and bridge to a video with a statement: "Although we have covered the working aspects of AI, there are meaningful and fundamental aspects that we are going to tackle today. Let's start off with a video..."	Before Class *Can we make consciousness into an engineering problem? | Aeon Essays*. (n.d.). Aeon. Retrieved November 30, 2021, from https://aeon.co/essays/can-we-make-consciousness-into-an-engineering-problem *Minds, brains, and programs | Behavioral and Brain Sciences | Cambridge Core*. (n.d.). Retrieved November 30, 2021, from https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/minds-brains-and-programs/DC644B47A4299C637C89772FACC2706A In Class

	- Showing short 30-second to 1-minute-long excerpts from two videos of human-AI interaction - FIRST: Will Smith talks to Sophia Will Smith. (2018, March 29). *Will Smith Tries Online Dating*. https://www.youtube.com/watch?v=Ml9v3wHLuWI - - *Guiding questions about the WS video:* - *"Was WS on a date with a person?" "Why/why not? What was missing?"* - SECOND: *Her movie video clips (excerpts) - Her—The beach scene*. (2014, November 3). https://www.youtube.com/watch?v=S7t4r2G2XCE - *Her—"Why do I love you?" scene*. (2016, May 4). https://www.youtube.com/watch?v=mWL9DeRrmfc - - Guiding questions about the Her clip: - Is Samantha a real girlfriend? Why/why not? - (How) Is Samantha intelligent? - Can Samantha feel?	*Her—The beach scene*. (2014, November 3). https://www.youtube.com/watch?v=S7t4r2G2XCE *Her—"Why do I love you?" scene*. (2016, May 4). https://www.youtube.com/watch?v=mWL9DeRrmfc Will Smith. (2018, March 29). *Will Smith Tries Online Dating*. https://www.youtube.com/watch?v=Ml9v3wHLuWI *While we are discussing the WS video + Her clip, we are taking notes on the whiteboard of the most important concepts they mention, connecting them and creating a mental map of the related concepts that we can refer back to throughout the session*
40	**Part 2 - Terminology and definitions** - **MAIN QUESTION: How can you know if something is person or AI? How can**	Before Class They read *Minds, brains, and programs | Behavioral and Brain Sciences | Cambridge Core*. (n.d.). Retrieved

	we tell if we are speaking with a Chinese room or a person? - Guiding questions: - What do we mean by `mind'? - What do we mean by `function'. - Does a machine think? How does it process information? - What is intelligence & artificial intelligence? - Is it possible to distinguish between a human and an artificially intelligent agent?	November 30, 2021, from https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/minds-brains-and-programs/DC644B47A4299C637C89772FACC2706A We are using the methodology of the *Turing Test*, but not explicitly mentioning it so that students are not overwhelmed with terminology. *Unpacking the required reading together – using the questions they had to submit before class.* *Connecting these clarificatory definitions with questions and notions that have been already raised in part 1.*
40	**Part 3 - The future of minds. Future technology and whether machines and technology can ever attain the status of mind.** **3.1 Stitching concepts and up-to-date examples** - We refer to some of the concepts on the whiteboard (from Part 2), such as mind, function, body, cognition etc. For each of these examples, we show a short clip, a picture or we share a story about contemporary AI. For example: "If you thought that being good at cognitive tasks is enough to deem an AI intelligent, look at these AIs beating humans at various games (show AI chess, AI Dota 2, AI GO). Would you say these are intelligent?	Drawing upon the theoretical issues we discussed in part 2, how does this inform the future direction of the practice of developing artificial intelligence. Before Class Boston Dynamics. (2021, August 17). *Atlas | Partners in Parkour*. https://www.youtube.com/watch?v=tF4DML7FIWk In Class Short excerpts from: Boston Dynamics. (2021, August 17). *Atlas | Partners in Parkour*. https://www.youtube.com/watch?v=tF4DML7FIWk Fredrik Knudsen. (2020, October 30). *Deep Blue | Down the Rabbit Hole*.

	We go back and forth between concepts and contemporary technological examples in this stitching manner **3.2 Critical questions about the future of AI (AI in "principle"). Questions in class to reflect on we have learned.** - Do you think that we will ever attain human level AI? If yes, when, and why? If no, why? - Show a survey from current AI experts which tells if they think AI will attain general human intelligence and when. This should serve a point that even experts do not have a consensus. - Can computers have minds in principle (theoretical)? - Can computers have minds in practice (engineering)? - What would we need to think about when creating computer minds? ***3.3 Final review/wrap-up of the lecture. Review of what we covered and learned in class thus far. Emphasize the most important aspects, linking it to our Learning Outcomes and also allude to the students what is coming next in the following session:*** - *Through revisiting our initial example:* - *Summarize functional theory of mind and its critics.* - *Plausibility of functionalism for AI development.* - *Current state of AI development, and future prospects.* **3.4. *Post class poll + one of the potential homework questions as essay themes:*** 1. *Can a person ever have a meaningful connection with an AI?*	https://www.youtube.com/watch?v=HwF229U2ba8 *The Surreal Dreams of AI-Generated Art | Video Essay*. (2021, October 17). https://www.youtube.com/watch?v=Bi4sJEE8wCs Two Minute Papers. (2019a, February 6). *DeepMind's AlphaStar Beats Humans 10-0 (or 1)*. https://www.youtube.com/watch?v=DMXvkbAtHNY Two Minute Papers. (2019b, May 18). *OpenAI Five Beats World Champion DOTA2 Team 2-0!* 🤘. https://www.youtube.com/watch?v=tfb6aEUMC04 Voy Zan. (2020, August 31). *AI Dungeon—This AI left me SPEECHLESS*. https://www.youtube.com/watch?v=09wVAUwm6bw

	2. *Can we ever know for certain if something is AI or natural intelligence?* 3. *Is human-level AI possible?* 4. *"What was most intriguing for you during this session?"*	

References

Boston Dynamics. (2021, August 17). *Atlas | Partners in Parkour*. https://www.youtube.com/watch?v=tF4DML7FIWk

Can we make consciousness into an engineering problem? | Aeon Essays. (n.d.). Aeon. Retrieved November 30, 2021, from https://aeon.co/essays/can-we-make-consciousness-into-an-engineering-problem

Fredrik Knudsen. (2020, October 30). *Deep Blue | Down the Rabbit Hole*. https://www.youtube.com/watch?v=HwF229U2ba8

Her—The beach scene. (2014, November 3). https://www.youtube.com/watch?v=S7t4r2G2XCE

Her—"Why do I love you?" scene. (2016, May 4). https://www.youtube.com/watch?v=mWL9DeRrmfc

Minds, brains, and programs | Behavioral and Brain Sciences | Cambridge Core. (n.d.). Retrieved November 30, 2021, from https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/minds-brains-and-programs/DC644B47A4299C637C89772FACC2706A

The Surreal Dreams of AI-Generated Art | Video Essay. (2021, October 17). https://www.youtube.com/watch?v=Bi4sJEE8wCs

Two Minute Papers. (2019a, February 6). *DeepMind's AlphaStar Beats Humans 10-0 (or 1)*. https://www.youtube.com/watch?v=DMXvkbAtHNY

Two Minute Papers. (2019b, May 18). *OpenAI Five Beats World Champion DOTA2 Team 2-0!* 😈. https://www.youtube.com/watch?v=tfb6aEUMC04

Voy Zan. (2020, August 31). *AI Dungeon—This AI left me SPEECHLESS*. https://www.youtube.com/watch?v=09wVAUwm6bw

Will Smith. (2018, March 29). *Will Smith Tries Online Dating*. https://www.youtube.com/watch?v=Ml9v3wHLuWI