Take-Home Synthesis Examination

UNIV 3275 Spring 2019: The Ethics of Artificial Intelligence
Available on Friday, March 8
Due Friday, March 29 at 11:59 pm to Brightspace

Examination Format

You have three questions, each formatted roughly as follows.
- You are asked to do close reading/viewing of two primary sources from required sources across two different weeks of the semester. The basic assumption of course is that you would have read them earlier during the respective weeks, but if not, this is an opportunity to catch up. Read closely and carefully.
- The questions will each ask you to
  - Write a formal essay as your answer to the question using the following guidelines:
    - No less than 1000 words and no more than 2000 words in length per question, not including references and acknowledgement in the word count,
    - using 12 pt Garamond font, 1.5 line spacing, and 1" margins,
    - APA format for references and citations,
    - submitted as a PDF;
  - reference and discuss the primary sources in a comparative manner (i.e., compare and contrast and synthesize, as appropriate);

- reference and fold in to your answer secondary sources as specified for each question. These secondary sources need not be discussed at length, but they should not be cited flippantly either -- their inclusion should be on point (to support your claim);
- include other references based on your own research, as you see fit.

It's tempting to start with the primary readings in order to define your essay themes, then work to include the secondary readings, specified by us or not. And this will probably be what many do on most questions, but that need not be the way you proceed, necessarily. You could start with a favorite theme and work to include the readings around it. See Grading, however, and make sure that the primary readings are given significant treatment as the main target of analysis.

Examination Themes

A meta-theme of this exam, which guides the questions, is *sublating binaries.*
In class, discussions and virtual forum posts have often exhibited a tendency toward "binarism," whereby two scenarios or ideas are framed as an "either/or" choice (aka *exclusive disjunction*). For instance, when exploring the present status and future prospects of incorporating AI into education, finance, sales, labor, and/or the legal domain, many have expressed concerns about AIs "replacing" humans or perhaps human functions, based on the notion that humans and AIs each have unique and at times even mutually exclusive characteristics, such as emotion and empathy (human) vs. speed, scale, and/or accuracy (AI). While binary or dichotomous analysis can be useful for charting strengths/weaknesses or benefits/drawbacks in view of constitutive AI applications, they can also deteriorate into identitarian politics or an essentialist ontology (gender, race, class, merit/intelligence, physical or mental 'fitness,' and more), the dangers of which we have discussed at length during the AI & Otherness week. Such an approach can readily lend itself to absolutist ethics whereby discrete modes of existence and intelligence get mapped on to hierarchies such as superior/inferior, or wrong/right. In writing your essay for each question, you should be mindful of these binaries with an eye toward actively challenging, qualifying, eliminating, or perhaps recuperating them as you see appropriate, making sure that your responses are informed by supporting sources and rhetorical efficacy.

Another meta-theme is ethics (of AI), of which we don't want you to lose sight. Ethical issues can be front and center, but if not, they should at least have a significant profile in the essay.

Each question has a subtheme, which are "*Perceptions of AIs, Humans, and Others;*" "*Assistive AIs;*" and "*The Nature of Intelligence*". These themes are not mutually exclusive. Surely you have noted that the materials across the course overlap considerably in the concepts that are addressed. Thus, it won't be a surprise that materials that are relevant for one question will be so for others as well. You cannot

change the required primary and secondary sources in a question, but you are free to address a primary source in one question as a secondary source in another question. You can also use the same source to satisfy the secondary source specifications of two questions (i.e., a secondary source can double count across questions).

Examination Questions

There are some materials that we do not give as primary or secondary sources (e.g., Race, Gender, & Power: a new agenda for AI research; the machine learning TA), because students thoroughly engaged with the material in class and in the VF. Your essays can certainly include these materials (or any others for that matter) as secondary sources that are integral to your themes.

1. Perceptions of AIs, Humans, and Others

This question asks you to
- identify contexts where the perceived distinctions between AIs, humans, and potentially others are blurred, porous, qualified, and/or non-existent; and
- reflect on how such contexts push us to reassess the ethical implications of how we regard and interact with artificial intelligences, and humans as well, based on perceptions and beliefs.

By "ethical implications," we are referring to values that extend beyond the good/bad, wrong/right, conscience/not, empathetic/not, and intelligent/not binaries. You can also address contexts in which perceived human/AI distinctions are reified or otherwise strengthened, perhaps in only particular circumstances, which is what we mean by 'qualified' above.

We ask you to focus on two primary sources in your analysis:
- "Issues of Authenticity in Autonomously Creative Systems," and
- Ted Chiang's novella *The Lifecycle of Software Objects*.

Compare, contrast, and otherwise analyze (e.g., how do the source materials differ and/or intersect on issues) and synthesize the two sources, and advance your claims through close readings/viewings of specific passages and/or scenes/sequences.

For secondary sources, your answers should draw on insights from
- at least one source selected from either the *AI & Personhood* week, or *AI & Otherness* week, or the Uncanny reading(s) for week 11; and
- "Legal Personhood for Artificial Intelligence" up to and including Section III (Could an Artificial Intelligence Serve as a Trustee), subsection A (The Scenario)

In this question we are asking about how perceptions of intelligence, empathy, consciousness, and creativity alter boundaries between humans, AIs, and others.

2. Assistive AIs

This question asks you to discuss how uncertainties, typically expressed as probabilities and expected utilities, typically expressed as weighted averages of "benefits" and "penalties," serve to benefit AI decision making as well as the decision making of humans that use AI tools, and perhaps collaborative human and AI collectives. While this question refers to probabilities and utilities, don't freak out, we don't expect mathematical answers. Your intuition should suffice.

The two primary readings for this question are
- "Smart Cities: Utopian Vision, Dystopian Reality," and
- "Computational Sustainability: Computational Methods for a Sustainable Environment, Economy, and Society"

Compare, contrast, and otherwise analyze (e.g., how do the source materials differ and/or intersect on issues) and synthesize the two sources, and advance your claims on ethical implications through close readings. In particular, identify areas of (potential) uncertainty and utility in the domains addressed in each reading, and methodologies and tools that might be shared across the domains of each reading, the role of uncertainty (or not) in each case, and what ethical questions are implied and addressed.

For secondary sources, include
- one or both 60 minute stories (2010, 2014) on high frequency trading; AND
- vignette 7 and/or vignette 8 from Professor Michael Bess' book chapter (Plan B); AND
- at least one additional secondary reading/video, which need not have been done for class, but could be.

3. The Nature of Intelligence

This question asks you to consider human, artificial, and collective intelligence as currently understood and realized, as well as imagined intelligences (e.g., the bicameral mind) and reflect, to the extent that you can, on the *nature* of intelligence.

The two primary readings for this question are
- *Neuromancer*, and
- the "If-Then-Else" episode from *Person of Interest*.

Compare, contrast, and otherwise analyze (e.g., how do the source materials differ and/or intersect on issues) and synthesize the two sources, and advance your claims on ethical implications through close readings. We don't expect you to be experts on human, artificial, and collective intelligence coming out of this class, but we do expect you to be able to look under the hood to a limited extent and not view various sources of intelligence as either mutually exclusive or as fundamentally the same. Address relationships between different sources of intelligence, and reflect, to the extent that you can, on the nature of intelligence. The focus of this question is to be much more about what you deem to be intrinsic qualities of intelligence rather than perceptions of human and AI distinctions and commonalities, which is the case in Question 1.

For secondary sources, include
- the Preface and Chapter 1 from N. Katherine Hayles's *Unthought: The Power of Cognitive Nonconscious*
- "From Design to Implementation to Practice a Learning by Teaching System: Betty's Brain"
- Under the Hood slides and concepts from Week 3 and 4.

Grading

There are baseline criteria on which we anticipate everyone to do very well. If you do a very good job on the following for all questions then you will receive a solid B (or 85/100, if you like).

The Baseline Criteria

- Satisfaction of specification (address all aspects of each question)
    - Reference and substantially discuss primary readings
    - Reference and include material from secondary readings
    - Adhere to specified format
- Integrity of writing (grammar, syntax, diction, style)
    - The best writing, in our experience, results from multiple editings and revisions
- Critical intervention and synthesis -- do your essays introduce new perspectives, insights, knowledge that is beyond your prior experience and insights found in the class material? There should be some of this in a solid B examination.
- The answer should address ethical issues and implications

Above and Beyond

To obtain better than a solid B, you must go above and beyond the basics. This admittedly involves subjective judgements on both your and our parts. We believe that your attitude in going above and beyond should be that you would be able to take risks (i.e., "thinking outside the box"), at least early in the reflection on a question -- we want this course to a safe space but also a *brave* space. However,

in the end you should ensure that your critical interventions and analyses hold together -- risks that result in disconnectedness suggest a last minute effort. This progression from creative, "adventurous" (or even perhaps "risky") early thoughts to a compelling essay often takes multiple passes and revisions, just as instructed in those aspects of the baseline criteria, but going above and beyond will often require reorganization along the way as well, or hefty reflection in advance of your first draft. We will be looking for things such as affective impact (on the part of the reader), intellectual ingenuity, and creativity.

We will also look at other characteristics of your answers as possibly above and beyond. We don't want to encourage finding additional sources that are simply cited flippantly (for the sake of fulfilling the citation requirement), but instead want you to locate and incorporate additional sources that you intelligently fold into your answers to go above and beyond. Note that such sources can include the VF posts, VF responses, TAs, of other student and faculty comments as well, which can indicate an engagement with the course that goes above and beyond. If you incorporate ideas from these latter kinds of sources, acknowledge them as footnotes or as citation to "personal communication." Remember, always acknowledge intellectual contributions by others -- it's one of the noblest aspects about the academia.

Suggestions

- Start early and work consistently. You need close reading/viewing, reflection, and question/discussion time after initial exposure to the questions and writing
- As you read/view materials from the primary and secondary sources for one question, take more of their possible relevance to other questions
- If you have questions, ask us -- we are happy to help if you want to run themes by us (and we know that you know you'll be assessed on the final product).

Our target deadline for having the graded exams back to you is April 16.

The Honor Code

The honor code for this examination allows you to talk to others and to treat them as any other source. You are obligated to acknowledge any ideas that you receive from them and use in your essay. You can point to influential discussions in an acknowledgment section or cite them as "personal communications." The choice is yours, but be consistent. Neither references or acknowledgments will count toward the word count.