**Artificial Intelligence Tech Talk: Transcription**
4/27/17

Tech Talk presented by Dr. Wayne Iba on 4/25/17

### *Slide #1: Humans, AI and Reality*

*Narration:*
If you took my class last semester, you know that AI is a huge and growing field, so there's no way that I'm going to give you *much* about AI

And then, of course, *reality;* that's an incidental topic that we can dive into.

### *Slide #2: Objectives*

- Motivate learning about AI
- Relax grip on current ideas and worldview
- Consider strange ideas and views

*Narration:*
Here are my objectives for this talk:

- I hope to *motivate* you to the point of learning more about AI.
- I also hope that you're going to *relax your grip* on the things that you know, or the things that you think you know; that's not to say that those things that you know or think you know turn out to be wrong, but I simply want you to relax your grip, because by entertaining strange ideas, we *improve* our thinking and learn to live in the world more effectively.

Those are my objectives.

### *Slide #3: Overview*

- What is AI and why should we care?
- Transhuman and posthuman
- Simulations and reality

*Narration:*
Here's an overview of where I hope to take us on this little journey.

First, I'll say a little bit about what AI is, and why we should care about what AI is.

Then, I'll end up talking about what simulations are and about the nature of reality.

In between, we're going to use transhumanism and posthumanism—and the singularity as a bridge between the two.

That's the sketch of the terrain that I want to explore.

I'm hoping that this will stimulate a lot of discussion. I'm aiming for 40 minutes; that should

allow plenty of time for questions. If you've got clarification questions or technical questions, we can take those as we go, but please hold the discussion, if we can, until the end.

***Slide #4: Why Know About AI?***

- Pervasive impacts on many facets of human life
- Analogies
  - Programming, cars and auto-mechanics
  - Net-neutrality
  - Copyright
- What policies should we adopt for AI?

*Narration:*
So why do I think that it's good for you to know about AI? Remember, I'm not promising to actually *tell you* about AI but I'm going to, hopefully, motivate you to learn more.

Well, as you know, AI is having a more and more pervasive impact on our lives. How many of you received a package this week from UPS or Fedex? [Several hands went up.] OK.

The efficiency of those delivery systems are made possible, in large part, by AI.

How many of you have taken some pharmaceutical medication this week [Several hands went up.]

A lot of drugs—not necessarily the ones you've taken—but AI is helping design drugs at rates that we wouldn't have thought of 50 years ago.

And how many of you arrived here today in a self-driving car? Actually, the fact is that none of you did, but you can see that's coming—it's not that far off.

And not all of the impacts are positive. Right?

We're facing a time where we may be losing jobs to AI. Or, losing our sense of self-worth or dignity; not just because of losing jobs but because we are no longer top dogs in the cognitive world.

And of course there's the ever-present fear of *The Terminator*, autonomous weapons, and so forth which again are appearing more and more on the horizon. I'm kind of shocked at the things that I read. It's almost like you get tired of being shocked and you're no longer shocked when you read, "oh, they actually built this autonomous weapon," or they meet robots that shoot guns.

So there are lots of good reasons to know about AI.

Consider: Humans commonly employ analogy to reason. And we particularly resort to analogy when we lack sufficient underlying technical understanding. So analogies are great in those cases but we should also recognize that they can steer us wrong. So we need to know that

we are using analogies—and that analogies are not proofs.

I'll give you a couple of examples:

*Programming:* I'm going to suggest that it's good for everyone to know how to program. You may or may not agree, but as a persuasive argument I'll say, "what about cars and auto mechanics?" Well we all drive cars—or at least most of us do—and many people who drive cars don't have any idea about the internal combustion engine and how the car works ... they know enough to stop at the gas station. So it might be hard to see why you would need to know about auto mechanics because cars have become so reliable.

Maybe a better analogy would be cars and taxis. All of *us* know how to drive because that's the world we grew up in, but if you grew up in a world of self-driving cars and you don't know how to drive, then you are at the mercy of where the car is going to take you. When you know how to drive, you can go where you want to go ... within bounds, abiding by the rules of the road. If you don't know how to drive, then you're relegated to the passenger seat.

Let's think about another example or another analogy: *Net-neutrality.*

There's not much hope of coming up with good policies unless we have some understanding of how the network works. And this is partly why we're in the position that we're in with a number of bad policies on a number of different issues. because the people making these policies don't have this underlying understanding, and they're not listening to advisors who *do* have that understanding.

And *copyright* might be a really good example because the introduction of a technology—and this is what we're talking about with AI—the introduction of a technology makes their old policies suspect. And being able to understand the technology that's in question will let us revise policies in sensible ways.

One question that we ought to be asking is: what policies should we adopt with regard to AI? Because AI is a tool and it's impacting many different aspects of our lives in new ways and there are policy vacuums there. We need to come up with understanding and in order to do that it would be helpful if we have some understanding of the technology.

AI has penetrated our popular media. I can affirm that all of you are familiar with AI at the movie level. There's a lot of leverage I can take advantage of—with your understanding.

***Slide #5: What is AI? (video)***
VIDEO: https://www.youtube.com/watch?v=etZPXWZpF7s

*Narration:*
That was kind of fun, for me. I don't know how many of you recognized these clips, including the original *Westworld.*

AI is penetrating our popular media and there's a certain amount of common understanding that we can affirm. And the story that overlays that montage—or mashup—of clips was kind of negative; there was a gloomy, mood-chilling effect and it doesn't need to be. Certainly that's the way it has penetrated our media in a number of cases.

***Slide #6: What is AI?***

- Computational artifacts that behave intelligently
- Search through a state-space
  - Start state
  - Goal state
  - Transitions
  - Search method and guidance

*Narration:*
That's the popular understanding of AI, so let's be a little bit more technical so I don't completely disappoint everyone.

We have defined AI as the attempt to construct computational artifacts that do things that seem to require intelligence. Another way to say that is: to perform tasks that, if a human were to do it—or perhaps *imagine* doing it—because ultimately, we would like to have AI systems that would accomplish tasks that we don't know how to solve but imagining solving those problems would require lots of thinking. That's one way to think about AI; we just want to build something that will do something that seems intelligent.

One of the most important pillars of AI is this notion of search through a state-space. In fact, we're kind of begging the question of: *what is intelligence?* When we say we want to build an artifact that behaves intelligently, we're not really saying what intelligence is. And this pillar, this foundation-stone of *search* really applies to intelligence.

We can say that *intelligence* amounts to search through a state-space, or work with state-space. You've got a state-space—you've got a bunch of states—you've got a Start-state, you've got a Goal-state, you've got transitions between those states.

Think about TicTacToe ... we're all familiar with TicTacToe ... there's a Start-state where the board is empty, and then there's a Goal-state where you've got three of your markers in a row. And there are transitions when it's X's turn; X can put a mark in one of the nine open squares, initially, and then, in response, O can put the O mark in one of the eight remaining squares, and so on. That defines a State-space.

We've got the Start-state, all of the states, all of the transitions between states—the *valid* transitions—so we can't go directly from here [first position] to here [last position] in one step because there are two moves involved. And then, the Goal-states are scattered throughout the State-space. And then, the objective of TicTacToe is to find a path through this State-space to a state where you've got three of your markers in a row. So that's one way of thinking about State-spaces.

Here's one example of State-space: this talk—this presentation—is the result of a search through a State-space; in a sense, it's a plan that I came up with in order to accomplish a set of objectives that I set. I said I wanted to motivate you to learn more about AI, I wanted you to relax your grip on certain, treasured pre-conceptions that you have, and I wanted you to entertain some strange ideas.

OK. Those are my goals; that defines the Goal-state. And then the Start-state is the combination of all of your current belief systems and knowledge states together with my imagination—or guess—at your respective knowledge and belief states; that's our Start-state. And then, the transitions are the things that I say and show and do and gestures I make and so forth. If I have an effective search then we'll transition through these states and we'll end up at a state where you are more motivated and you are relaxing your grip on preconceived ideas.

***Slide #7: Search and AI***

- Bayesian belief networks
- Neural networks
- Representation engineering
- Application engineering

*Narration:*
This search through State-space is incredibly fundamental and pervasive. Now, even when we get to some of the more relatively-newer and sexier AI techniques—Bayesian belief networks and artificial neural networks—search is still at work.

What we can think about a Bayesian belief network is: as a search through a space of conditional probability tables that reflect the interactions—or the inter-dependence—between events in particularly random variables and, for artificial neural networks, we can think of them as a search through the space of weight vectors; each one of the artificial neurons within a neural network as a collection of weights, and we can think of the entire network as being a vector of weights. We want to search through the space of vectors for a set of weights that will let the network compute certain results that correspond to solutions that we care about in the real world.

For all of this—and it's especially important with neural networks and Bayesian belief networks but even with classical search—most of it boils down to representation engineering. You've got to change the problem in the real world; you've got to come up with a *model* of the problem and represent that model in a way that is compatible with the technique or the tool that you're intending to apply, whether it's a Bayesian net or a neural net.

And, this is something that is kind of overlooked—and maybe under-appreciated—that maybe 90% of the effort necessary to build successful AI applications is representation engineering.

In a replication of that same principle, AI itself, employing AI tools to solve some problem is an example of application engineering. When I'm talking about application engineering, I'm

not talking about engineering an application ... like, if you've got something you want to accomplish, you will *engineer* a solution; you'll engineer an application that solves that problem. This is more the other way around; it's not engineering an application, it's application engineering. You're saying, "well, these are the tools that I have available, what are the applications that I can apply this tool to? ... not, "here's the problem that I want to solve, how do I build something to do that?" It's, "Here's my tool set, what can I do with that tool set? Is there an application there that I can solve?" ... and, if so, then maybe there's a product there, I don't know, but then we can do something. *That* I'm calling application engineering; building an application.

*Audience Question:* So the representation engineering is creating a model of the system that you want to do something with ... *[Wayne]:* plus taking that model and making a concrete representation that can be represented in a computer system ... *[Audience Member]:* and the application engineering is looking at tool boxes that might help the goal of the project? *[Wayne]:* Yes.

To try and make this analogy a little more parallel, representation engineering is taking a real world problem, working it into a model and then implementing that model—or instantiating that model. Application engineering then is saying you don't have exactly the same sort of model intermediary but you're taking the real world—maybe things that you want to accomplish—and you're constrained by the tools that are available, and how can I—or even, *can* I—solve that problem with the tools that are available to me. And, often times, it's cobbling-together multiple tools that are going to end up letting me solve the overall problem that I'm targeting. But many other times, I'm going to end up just saying, "well, no, I can't solve that just yet; we need some tools, or there's more work to be done before we can do that."

The distinction that I'm trying to make is: With regards to AI and how it is impacting our lives, it's not about a particular application and how we build a system to solve that; it's more about how we're going about taking AI tools and building AI tools to solve more and more problems.

### ***Slide #8: So What?***

- Augmented reality
- Singularity
- Simulations and reality

*Narration:*
As AI becomes more and more of a commodity and penetrates more and more areas of our lives, we're going to encounter more and more augmented reality. In fact, actually, we've had augmented reality forever. As soon as someone built a wheel, that was an augmented reality; it let us do things we couldn't do before. And the lever let us move things that we couldn't move with a whole bunch of people. And so on. But now, with AI, the augmentations are more and more significant and happening more and more rapidly, and people have talked about the singularity ... and, as we are more and more augmented and as AI continues to advance, we might reach a point where the intelligent creations that we are building are

more intelligent than we are, and start to create their own intelligent things.

Then it will take off and we're left in the dust; that's one characterization of the singularity. Even in hindsight—if we survive the singularity—we're not going to be able to pinpoint, oh, there it was. It's this transition time from human power cognition to computational power cognition.

As this transition is taking place, simulations are becoming more and more important. One of the techniques that's very popular in Artificial Intelligence is Monte Carlo Simulation; that's an important tool, but with Augmented Reality comes Virtual Reality and some of the cognitive work that's going to be taking place by our intelligent systems is going to take place in simulated worlds. They're not going to be thinking with *physical* blocks. Children learn a lot of cognitive skills by manipulating physical blocks. But a lot of the thinking work that AI systems are going to be doing is going to be taking place in virtual worlds, simulated worlds, imaginary worlds. So, I want us to think about simulations and reality. Because what we might find is: after reflecting on simulations and reality, which turn out to merge in this kind of strange space. We may find ourselves back at the beginning with maybe—or, hopefully—some insight into policies for AI, or how to make sense of Ai, and whether or not we should be pursuing AI.

***Slide #9: What Is Real?***
<Second Life simulated image of room showing two people sitting at front of room, with various projected images near them, and audience members seated in theater-style chairs>

*Narration:*
Here's another simulated image...How many of you have *not* seen *The Matrix?*
[Not one person raised their hand.]

***Slide #10: What Is Real?***
The Desert of the Real (video clip from The Matrix): https://www.youtube.com/watch?v=tgBViHeiSKM

Morpheus explains his (simulated) surroundings to Neo.

*Narration:*
I just love that! Artificial Intelligence. And Neo is, remember, this eminent hacker.

So we've got *Second Life* as one example, and *The Matrix* as another example of simulated worlds. Obviously, *The Matrix* looks a lot more like our world, right? Well, of course, that's because it was filmed by humans in the real world and was supposed to be a simulation. But never the less if we go along with the writer/author and play along while suspending disbelief maybe there's a simulation that has such a higher fidelity to the real world than *Second Life* has, maybe we can interact with it.

***Slide #11: What Is The Problem?***
- Underplay role of the simulation for 'brain in a vat' scenarios

- Simulated worlds are real worlds
  - Calling for conceptual re-analysis
- Implications of expanded definitions

*Narration:*
Too often the role of the simulation itself—the simulation proper—the system that is the simulation—in this case, *The Matrix*—is sort of obscured in these 'brain in a vat' scenarios. So *The Matrix* is a classic instance of the 'brain in a vat' thought experiment that authors have used at least since Descartes, and the notion is, since you've all seen *The Matrix*, a human is —or it doesn't even have to be a whole human, it could just be a brain—is plugged into a simulation and getting inputs and outputs with the simulation for plugging into into the simulation and the simulation is handling all the physics of the real world.

So the role of that simulation is underplayed. Because what we have, currently, that we can interact with in simulations is *Second Life* which doesn't look anything like our real world. But if we imagine *The Matrix* where it's just like this world, the simulation that's necessary in order to accomplish that is really huge; there's a lot of work that has to be done to reach that point. But actually, more importantly, it's a confusion—it's a conceptual muddle—that simulated worlds actually are real worlds. And because of that ... I'm going to try to persuade you of that in a little bit ... but because of that claim, it calls for a conceptual re-analysis of what it means to be real and what it means to be simulated. And then, based on that re-analysis, we're going to have some implications for the thoughts that we hold near-and-dear.

### *Slide #12: Ontological Status of Simulations*
- Simulations are clearly real in the physical
- Simulated objects are real within simulation
- Reluctance to say simulated and physical objects both real

*Narration:*
I'm going to say—and I think you'll agree—that simulations are clearly real in the world. If I give you a simulation, you're going to say, "yes, you gave me something real, I can work it on my computer; it's actually something you gave me." Right? So, simulations are clearly real in the physical world. And I'm going to say—and I think you'll probably agree with me on this also—that simulated objects within the simulation are also real within the simulation. So if we had a simulated humanoid and a simulated chair existing within our simulation, the simulated humanoid could sit in the simulated chair and the simulated chair would support the simulated humanoid's weight. Unless the chair was made of paper and the simulated humanoid had a very high simulated mass, maybe the chair would collapse. But the point is: that the simulated object has a reality within the simulation. So I think you can agree with me on that point. And, if you don't, we can discuss that later.

I feel like there is a natural and intuitive reluctance to say that simulated objects and physical objects are both real in the same way. And so, I'm going to focus on that for a second.

### *Slide #13: Is This a Simulation?*
- Nick Bostrom suggests that we might be living in a simulation

- Not catastrophic if we grant the ontological status of simulated worlds
- What about implications for our worldview?

*Narration:*
Nick Bostrom is a philosopher in England. He, among other people, has suggested that we might be living in a simulation. So the physical world that we take for granted every morning could be a simulation. The details of his argument aren't relevant for our conversation here, but I'm going to appeal to him as a semi-authority and say, "some people are entertaining this question—or, entertaining this notion." And, with his analysis, it comes up to: lacking a better prior probability distribution over certain variables in his model, he ends up saying, "well, there's a one-third chance that we're living in a simulation, right now." So we could argue about the quality of his model but it's not completely out of the question that we are living in a simulation.

*Audience Question:* As AI grows, does this model have a better probability of becoming fact?

*Wayne:* No, I don't think that. Whether we are conceptual or not isn't going to have an impact on his probability. If we're never successful, we just happen to be a leaf in the hierarchy of the simulation.

***Slide #14: Exercise: What If?***

- Suppose we live in a simulation
  - What is real or not?
  - What does it mean to be human?

*Narration:*
I'm going to ask you to kind of relax your grip on your near-and-dear thoughts. Suppose we *are* in a simulation ... I'm going to say that it's not catastrophic if we grant the ontological status of simulated worlds. It strikes me as funny that the new movie, *Arrival*—and, how about *Contact*—in both of those movies, our sense of reality gets turned upside down because we are confronted with an alien species and many of the things we believe to be true are no longer true. And, in both of these movies, there are cults that perform mass suicide because they cannot take this revelation. But I'm going to say that that was a bad move. And in fact when it comes to simulations, if it turns out that we really are simulated—we were given proof of that—that it wouldn't call for these kinds of mass suicides. In fact, there are going to be some things, some implications of that but I think we can work through them.

So, work with me on this exercise. What if ... suppose we are in a simulation, we need to revise what's real and what's not. In actuality, the chair [points to a chair], that chair right there is as real as it was before I thought that it was simulated. It's still going to support my weight if I sit in it. If I kick it, it's going to hurt my foot. And if I inspect what it's made of, with a microscope or other tool, I'm going to find that it's made of similar things that I'm made of. It's still real. And, perhaps a more challenging question, what does it now mean to human?

Again, I'm going to say, "I'm still thinking the same way I was thinking before I found out that this was a simulation. I still have the same drive, the same goals, the same values. Basically, nothing has changed except this sort of underlying assumption of what it is that we're really made of." If we created a simulation where we had an intelligent simulated humanoid and it started asking this question and it took this simulated chair and started inspecting what it's made of and found that it's made of the same stuff, it wouldn't really matter if it finds that it's made of particles because we've simulated particles analogous to the particles in our world, or whether it discovers that the chair is made of one's and zero's. It doesn't matter because it's going to find that the humanoid itself is made of one's and zero's. So the constitution of the simulated universe that the humanoid lives in is staying the same; it's just a gawking new revelation. Think of it, by analogy, to when the Theory of Phlogiston was replaced by oxygen and elements. Phlogiston was this hypothetical fire material ... but then, we said, "No, phlogiston doesn't exist; things are not made of phlogiston and other stuff, it's oxygen and carbon and whatever." So, I'm going to say that it's not catastrophic.

### *Slide #15: Exercise Results*

- Three possibilities:
  - No problem
  - No possible way to reconcile
  - Theory revision

*Narration:*
Now, as I see it:

1. You might say, "yeah, it's working; no big deal." I'm not ready to go that far. I mean, maybe you're not paying attention or you're not awake if you say that.
2. You might also say, "there's no possible way to reconcile this; my life is over." I think that's a mistake.
3. Instead, we've got something that's analogous to the Theory Revision that was necessary when we discovered that phlogiston wasn't a basic stuff. Or, when we discovered in the Copernican Revolution that the Sun doesn't revolve around the Earth but instead, the Earth revolves around the Sun. *Joke:* Actually, the Sun *does* revolve around the Earth; it's just that thinking about it the other way makes the math come out a lot easier.

*Audience Participation:* "A relative joke..."

### *Slide #16: Discussion: AI*

- What do we gain?
  - A tool both similar and different
  - Extends/augments our abilities
- What do we lose?
  - Raises issues of autonomy, sovereignty, and humanity

*Narration:*
If we continue to pursue AI and happen to be successful, what do we gain?

AI is a tool that is both similar to and different from other tools that we have. It's similar to it insofar as it augments our ability. It's different in terms of the degree to which it's augmenting our ability and the pace at which it is doing so.

And what do we lose if we are successful? It does raise some serious issues. We like to think of ourselves as sovereign—or semi-sovereign—individuals, with free will and so forth. And if we are successful at building autonomous intelligent agents that are every bit as intelligent as we are, we need to ask questions about its autonomy, its sovereignty and, once again, what does it mean to be human if now there's this other thing that's operating much like we operate, then what does that mean for us?

### ***Slide #17: Discussion: Simulations & Reality***

- What do we gain?
  - Better precision with words/concepts: real
  - Plausible response if we live in a simulation
- What do we lose?
  - ???
- What else?
  - Possible guidance on ethics of simulations

*Narration:*
We can think about simulations and reality in a similar kind of way. If we accept this notion that simulations *are* real and reality *might* be simulated, what do we gain? Well, I think we get a little bit better precision with the working concepts that we use, particularly with real, but also with simulations. And also, we get a plausible response if it turns out that we *are* living in a simulation.

What do we lose? I'm not sure that we lose anything of terrific significance and it actually also gives us possible guidance on how to go about formulating ethical theories on how to running simulations. So is it right—or is it a good thing—to run a simulation? If we have simulated rats in a simulated maze, and we wouldn't perform certain experiments on real rats in a real maze, then is it right to perform experiments on simulated rats in a simulated maze?

I hope that motivated you to want to know more about AI, and I hope you kind of relaxed about things that you thought you knew. I look forward to discussions.

Transcribed By: Karen
Transcription Date: 4/27/17