AI, Ethics, & Critical Technical Practice

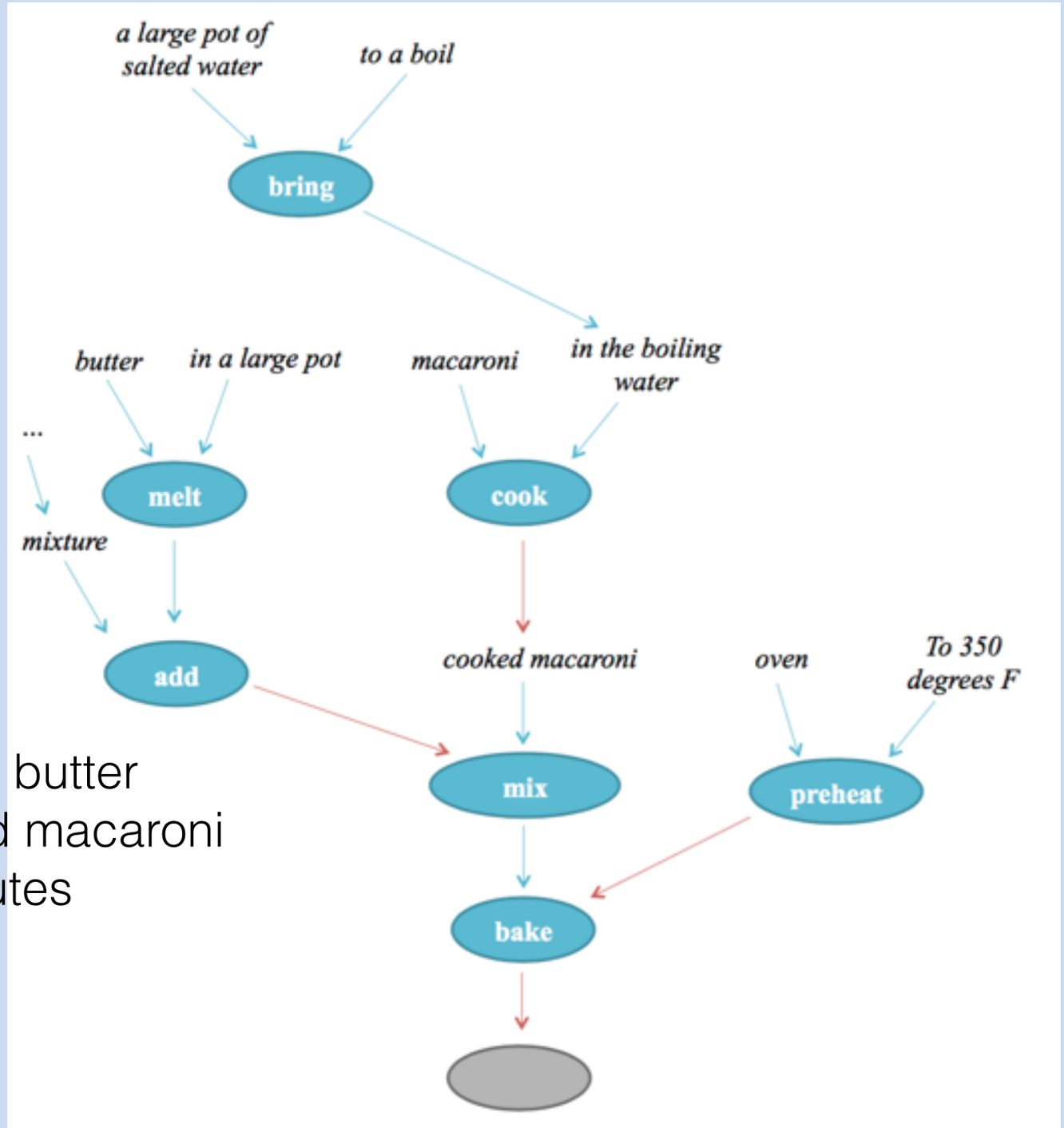
Knowledge Representation

- The computer can only do math
- We need to tell it what math to do
- We describe a system via a convenient *encoding*

Knowledge Representation

"Make macaroni and cheese"

- preheat oven to 350° F
- add water to a pan
- boil water in the pan
- add pasta to boiling water
- cook pasta
- drain pasta
- add butter to a pan
- melt butter in the pan
- add cheese mixture to the pan with the butter
- mix cheese mixture, butter, and cooked macaroni
- put pan into oven and bake for 10 minutes



http://www3.cs.stonybrook.edu/~gponnuraj/thesis.pdf

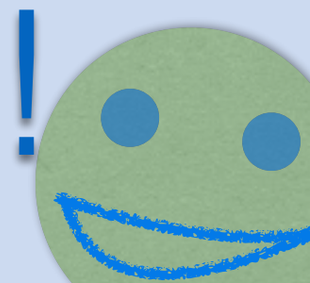
Knowledge Representation

- High level architecture
- Algorithm design
    - Behavior *and* inputs/outputs
- Feature design for machine learning
- Important for performance, efficiency, and even ethical concerns

Knowledge Representation

What knowledge representations
are used by AI systems you're
familiar with?

What are some benefits and
drawbacks of that representation?



People

- AI systems are *made for* people
  - Who are complicated, unpredictable, highly context-sensitive…
- Even worse:
  - AI systems are *made by* people!

People

- Are of a time, a place, a society, a class…
- Have particular (non-universal!) experiences
- Are *biased* explicitly and implicitly

People

- We are necessarily biased when we:
  - Pick a system to simulate/automate/AI-ify
    - Pick a knowledge representation
    - Implement/train/test/deploy the AI
- We can’t avoid our bias, but we must be aware of it

Fortunately we are smart people and have found a way out of this predicament. Instead of relying on algorithms, which we can be accused of manipulating for our benefit, we have turned to machine learning, an ingenious way of disclaiming responsibility for anything. Machine learning is like money laundering for bias. It's a clean, mathematical apparatus that gives the status quo the aura of logical inevitability. The numbers don't lie.

http://idlewords.com/talks/sase_panel.htm

Ethical Issues in AI

Do you know any examples of ethical problems in AI/computational systems/ algorithms/etc?

As a result, for the last 14 years, every time MaxMind's database has been queried about the location of an IP address in the United States it can't identify, it has spit out the default location of a spot two hours away from the geographic center of the country. This happens a lot: 5,000 companies rely on MaxMind's IP mapping information, and in all, there are now over *600 million* IP addresses associated with that default coordinate. If any of those IP addresses are used by a scammer, or a computer thief, or a suicidal person contacting a help line, MaxMind's database places them at the same spot: 38.0000,-97.0000.

Which happens to be in the front yard of Joyce Taylor's house.

http://fusion.net/story/287592/internet-mapping-glitch-kansas-farm/

In 2011, there were studies suggesting that when people saw positive posts from friends on Facebook, it made them feel bad. We thought it was important to look into this, to see if this assertion was valid and to see if there was anything we should change about Facebook. Earlier this year, our own research was published, indicating that people respond positively to positive posts from their friends.

Although this subject matter was important to research, we were unprepared for the reaction the paper received when it was published and have taken to heart the comments and criticism. It is clear now that there are things we should have done differently. For example, we should have considered other non-experimental ways to do this research. The research would also have benefited from more extensive review by a wider and more senior group of people. Last, in releasing the study, we failed to communicate clearly why and how we did it.

http://newsroom.fb.com/news/2014/10/research-at-facebook/

Uber seems to offer better service in areas with more white people. That raises some tough questions

But algorithmic decision-making takes on a new level of significance when it moves beyond sifting your search results and into the realm of public policy. The algorithms that dominate policymaking—particularly in public services such as law enforcement, welfare, and child protection—act less like data sifters and more like gatekeepers, mediating access to public resources, assessing risks, and sorting groups of people into "deserving" and "undeserving" and "suspicious" and "unsuspicious" categories.

Nine states based their refusal to disclose details about their criminal justice algorithms on the claim that the information was really owned by a company. This implication is that releasing the algorithm would harm the firm that developed it. A common recidivism-risk questionnaire, called the LSI-R, turns out to be a commercial product, protected by copyright. States such as Hawaii and Maine claimed that prevented its disclosure to the public.

https://www.washingtonpost.com/news/wonk/wp/2016/03/10/uber-seems-to-offer-better-service-in-areas-with-more-white-people-that-raises-some-tough-questions/

http://www.slate.com/articles/technology/future_tense/2015/04/the_dangers_of_letting_algorithms_enforce_policy.html

http://bigstory.ap.org/article/1efa2f2f5f004295a35d175f58149e67/we-need-know-algorithms-government-uses-make-important

The logical place for us to engage with a massive group of users was Twitter. Unfortunately, in the first 24 hours of coming online, a coordinated attack by a subset of people exploited a vulnerability in Tay. Although we had prepared for many types of abuses of the system, we had made a critical oversight for this specific attack. As a result, Tay tweeted wildly inappropriate and reprehensible words and images. We take full responsibility for not seeing this possibility ahead of time. We will take this lesson forward as well as those from our experiences in China, Japan and the U.S. Right now, we are hard at work addressing the specific vulnerability that was exposed by the attack on Tay.

http://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/#sm.001t2aaqz15nld0zvdg27tweevfg0

teënage shirtbag
@UnburntWitch

Follow

It's 2016. If you're not asking yourself "how could this be used to hurt someone" in your design/engineering process, you've failed.

Phew

What can you do?

- “Look outside of computing” as Harrell suggests
  - and *empower users*
- Broaden your perspective
- Become aware of biases
  - Intuitions are often simplistic or wrong
- Build diverse project teams
- Be humble

Critical Technical Practice

"…in which rigorous reflection upon technical ideas and practices becomes an integral part of day-to-day technical work itself."

Critical Technical Practice

“When technical work stalls, practical reality is trying to tell us something.”

“Mentalism”

- Computer ~ Brain
- Computation ~ Cognition
- Descartes via Turing: *Input/output problems*

Planning

- Agents execute hierarchical *plans* to solve *problems*

- But: what about reactions and improvisation?

Planning

Activity is organized by plans

Contingency is marginal

The world tries to foil plans, so plan extensively to avoid difficulty

The interactions of everyday life are orderly and patterned

Activity is constantly improvised

Life is almost wholly routine, the environment supports the activity

Metaphor systems

- These viewpoints are all *choices* we make, often without realizing it, when building software:
  - Information as a commodity
  - Society as a network
  - Formalization as hygiene
  - Computation as power