concept map

language model behavior

- observations
    - sycophancy (friendly; they aim to please)
    - hallucinations: confident in incorrect information
        - often don't realize or admit they are wrong
        - needs to be checked
    - "thinking" is less unique than human thinking
    - prompt hacking: ways around 'failsafes'
    - links to web pages (sometimes real, sometimes less trustworthy or outdated sites)
    - gpt-5 mimics human speech very well including hesitation, breathing, and stuttering
    - asymptotic improvement: diminishing improvement of each generation of models
        - GPT5 not doing the best
    - chatgpt asks for preferred responses out of two options
    - many different companies are creating "their own" llms for employees
    - new models introduced regularly
    - good at coding, not debugging
    - you can give them access to your computer to perform tasks
    - agentic AI
    - recursively trained models
    - unhumanlike learning (owls from numbers)
    - Grok has shifted to be more "right-wing"
    - LLMs memorize keywords and adjust settings to that word in later responsese
    - AI grows and learns like humans do (GPT5 learned from GPT4)
    - can produce bias like political propaganda or harmful content
    - really good at synthesizing info from the web
    - conveys nuance and subtlety
    - getting better at limited tasks in a small environment
    - has its own writing style (and not yet that of the average human)
    - reasoning ability
    - vibe coding: Claude Code writes entire apps based on instruction
    - LLMS are fairly similar
        - variation comes from the user
        - ChatGPT adapts to user style
    - no new concepts; just reminds us what we know

- questions
    - How will we be able to distinguish human and AI language?
    - Can AI have emotions? How important is emotional intelligence?
    - Do LLMs have a "plateau" where they can no longer be improved?
    - How do human inputs affect future outputs?

        - How do companies personalize LLM responses?
    - When will we have AGI?
    - How do LLMs "visualize" linguistic principles?
    - How are models trained to have a particular bias?
    - What's the consequence of training AI using AI/?

human interaction / uses / functions of AI

- observations
    - multiple cases of llms encouraging suicide and self harm; AI-induced psychosis
    - people use Character AI for relatoinships
    - alternative to therapy
    - art and design
    - work in corporate jobs
    - flood of AI generated spam/content
        - "AI ASMR videos are tuff"
        - social media trends using AI
        - "brainrot"
    - growing use in educational contexts; cheating in school
    - increasing reliance/overreliance on AI
    - AI voice bots on phone calls
    - good for learning
    - recognizing AI content is getting harder
        - especially challenging for older generations
    - renders "Excel people" useless
    - the AI boom is in full swing
        - many startups that are 'wrappers' around AI
    - intentions to automate as much as possible
    - AI (not LLMS) for driving, music recommendations
- questions
    - How will cheating with AI affect schooling? How will AI affect education in general?
    - How will AI affect mental health?
        - How can we prevent harmful emotional relationships between AI and humans?
    - How will AI change the way software engineers work?
    - Is there an overuse or over-reliance on LLMs?
    - Is AI replacing entry level jobs?
    - environmental impact?
    - What is the status of gov. regulation for AI?
    - How can we limit AI content from filling our spaces?
    - How does interacting with models compare to interacting with humans?
    - Does it change the way we speak?

attitudes about AI

- observations (our own opinions and others)
    - politically divisive
    - particular news outlets tend to hype or glorify LLMs
    - some people have blind trust in LLMs
    - risks and drawbacks not a large component of public discourse
    - "the LLM bubble will burst"
    - anxieties about labor market
        - AI shouldnt replace artists
        - CS grads experiencing uncertainty about the job market
    - confidence about labor marker
        - UT professors don't think AI will take CS jobs
        - people who prepare (by using LLMs now) will be bettter off
    - "we'll end up in a dystopia"
        - sooner or later models will be used for more bad than good
        - fear and anxiety about reasoning abilities and self-awareness
        - apocalyptical fears
    - "we'll end up in a utopia"
        - some think that AI will lead to an age of abundance will lead to the establishment of universal basic income (UBI)
- questions
    - Will the LLM bubble burst? (like how voice assistants did?)
    - Why train the model to respond in a specific register?
    - Why replace humans?

data

- observations
    - regulation is a route to privacy and security
    - chatgpt knows personal information if you typed it in earlier
- questions
    - How much of the internet should AI have access to? Is it even possible to limit what LLM's can access online?
    - How private is AI? Is my data cooked?
    - When does my data get erased?
    - How can we use LLMs without our personal data and chats being collected?
    - How do you tailor the internet data that you feed the model?
    - Where do AI models source their information?
    - Will people begin to think with AI rather than with themselves?

economy / corporations / governance?

- observations
    - companies prioritize engagement and profit over morality or user wellbeing

    - Talent war over AI researchers (Meta and OpenAI offering massive signing bonuses)
    - AI alignment
    - AI arms race
- questions
    - How can we make companies accountable? for their user's safety
    - How can we help against job losses?
    - What are the impacts on job growth / unemployment?
    - How can we slow down (development of AI(and ensure people are okay?
    - How are companies training AI models for their own benefit?
    - What are social economic political consequences of "misaligned" AI?
    - How to mitigate profit incentive?

relevant parties

- indiviuduals
- publics
- the law / legal system / government
- companies and corporations
- AI companies