CPSC 436C Cloud Computing

Winter 2025 Term 1 (September 9, 2025)

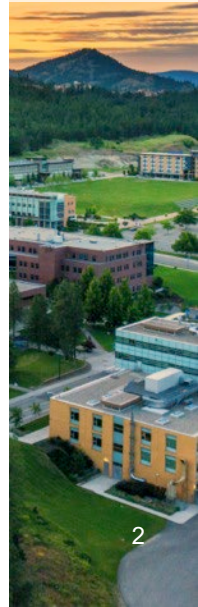
Tony Mason (fsgeek@cs.ubc.ca), Lecturer

Logistics

Tony's Office Hours:

- **Monday September 15 @ 16:00-17:00 (Discord)**
- **One more TBA**

TA Office Hours: TBA



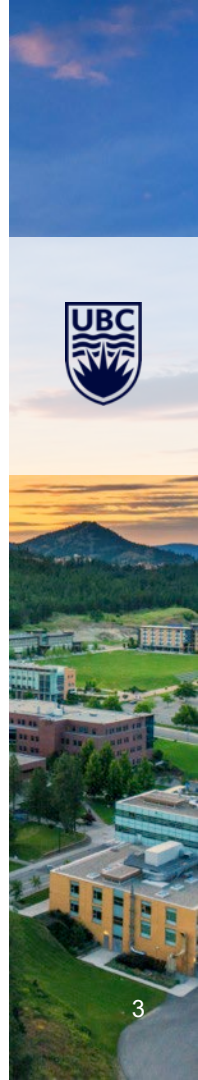
Resources

Observation: You need more pre-requisite understanding

- First example is a link shortener
- Requires you have cloud resources
- Benefits from AI mechanisms

Lecture structured:

- Discuss cloud resource options
- Discuss AI options
- Introduce (Part 1) of the link shortener



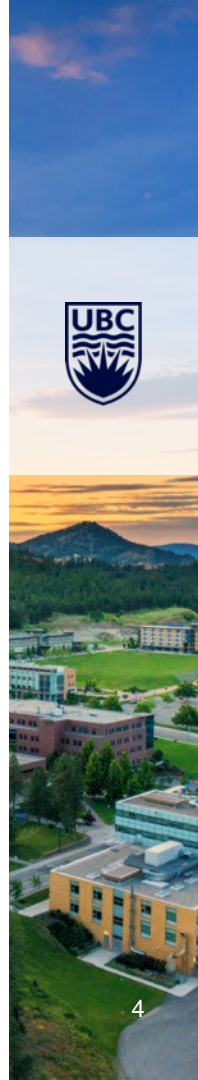
Job Interview Reality Check

Scene: *you are in a job interview*

Use AI to help solve this architecture problem

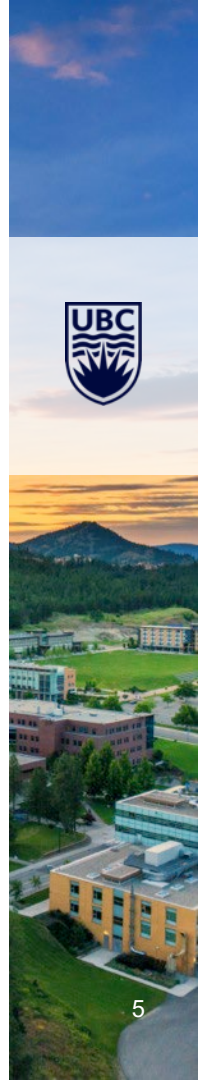
Question: what distinguishes hired candidates from rejected ones?

Answer: How you collaborate with AI, not just use it matters



Sources

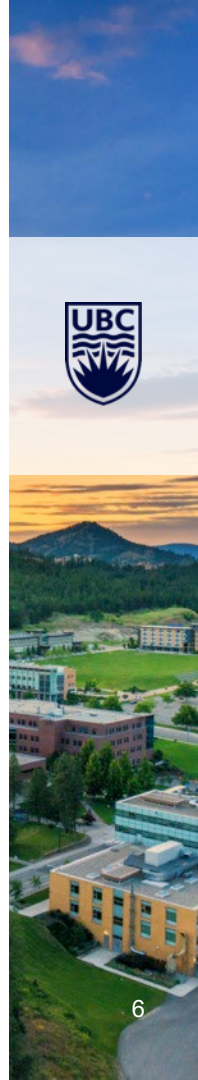
https://www.nature.com/articles/s41562-024-02024-1
https://www.upskillist.com/blog/7-ways-ai-boosts-creative-problem-solving/
https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.21.010149
https://news.mit.edu/2023/multi-ai-collaboration-helps-reasoning-factual-accuracy-language-models-0918
https://www.finalroundai.com/blog/what-are-the-benefits-of-using-ai-for-interview-preparation-a-comprehensive-overview
https://www.tandfonline.com/doi/full/10.1080/14703297.2025.2539779?src=
https://www.atlassian.com/blog/productivity/ai-collaboration-report
https://pmc.ncbi.nlm.nih.gov/articles/PMC12292075/
https://mitsloan.mit.edu/ideas-made-to-matter/when-humans-and-ai-work-best-together-and-when-each-better-alone
https://www.nature.com/articles/s41598-025-98385-2
https://www.tandfonline.com/doi/full/10.1080/08874417.2025.2492885
https://vidcruiter.com/interview/intelligence/ai-interviews/
https://pmc.ncbi.nlm.nih.gov/articles/PMC10643528/
https://www.sciencedirect.com/science/article/pii/S0268401224001014
https://www.canada.ca/en/public-service-commission/services/appointment-framework/guides-tools-appointment-framework/ai-hiring-process.html
https://www.sciencedirect.com/science/article/pii/S2666920X23000176
https://www.hrreporter.com/focus-areas/automation-ai/canadian-employers-turning-to-ai-bots-for-job-interviews-report/392632
https://arxiv.org/html/2407.19098v1
https://www.forbes.com/sites/keithferrazzi/2025/03/27/the-ai-recruitment-takeover-redefining-hiring-in-the-digital-age/
https://smythos.com/developers/agent-development/human-ai-collaboration-research/



Demo: Let’s ask some questions to see this

Query: "Help me build a scalable API"

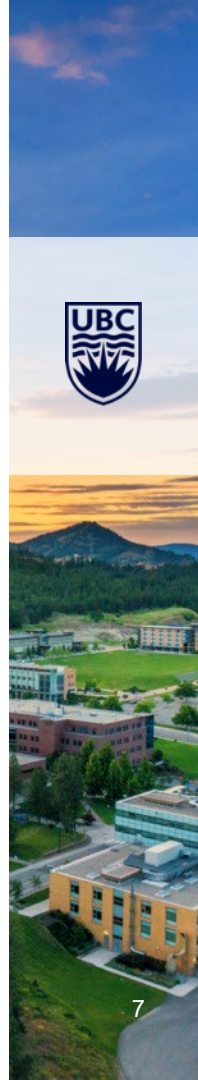
Query: “I need a prediction API that scales 100→50K req/hour, AWS free tier, <200ms latency. Compare serverless vs containers with cost breakdown”



Demo: Let’s get real

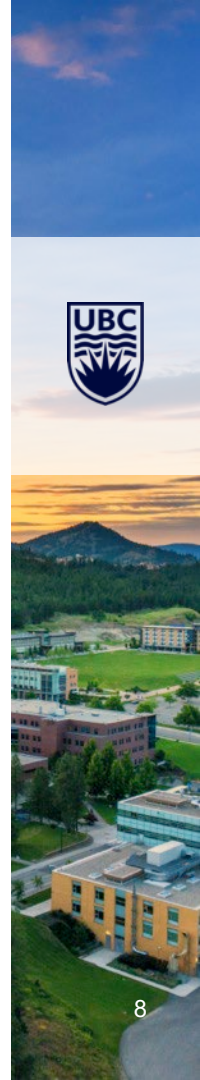
Architecture is one thing, but most of our time is spent inside the code:

- Looking up documentation
- Finding irrelevant or deprecated results
- Debugging code – our own, and that which we obtained elsewhere



Example Code

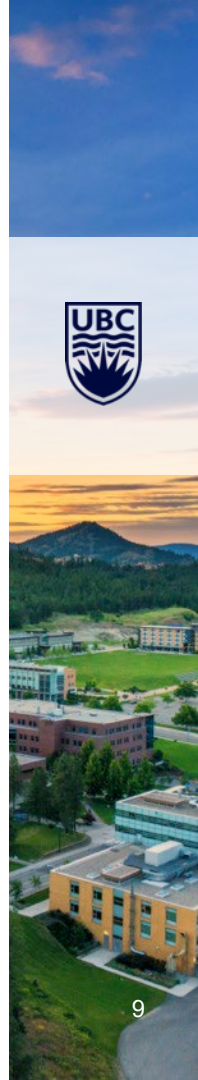
```
def parse_url(url):
    # A naive and broken URL parser
    parts = url.split('://')
    protocol = parts[0]
    rest = parts[1]

    site_path = rest.split('/', 1)
    site = site_path[0]
    path = site_path[1] if len(site_path) > 1 else '/'

    return {
        'protocol': protocol,
        'site': site,
        'path': path
    }
```



Debug Query

Query 1: *Why isn’t this Python code working?*

Query 2: *This Python function is meant to parse a URL and return its components, but it fails on URLs with encoded characters and fragments. The error is a KeyError on 'fragment'. Analyze the code, explain the logical error, and provide a fixed version.*



What Just Happened?

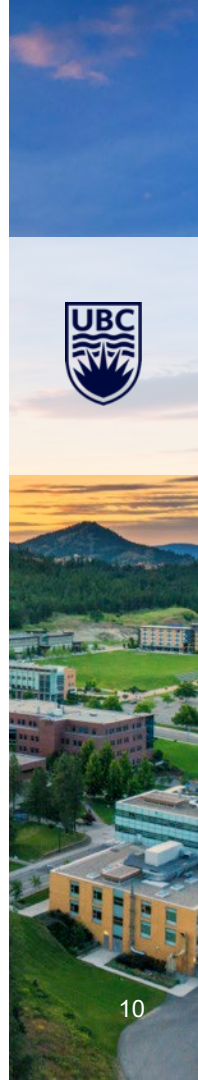
Constraints *improve* responses

Specific context gets actionable advice

Different models have different strengths

Your job: critical evaluation of AI output

If your classmates don’t understand your question, LLMs won’t either.



What you need

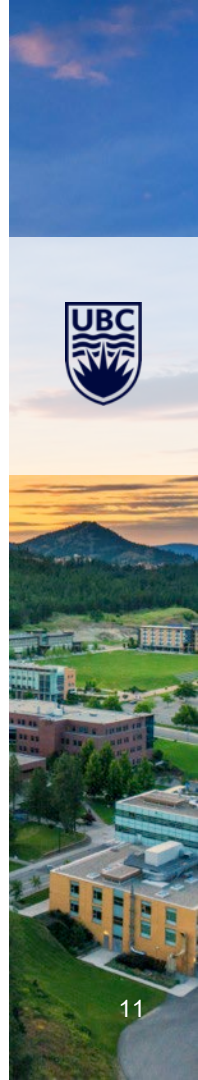
AI access

- Local (ollama, LM studio, etc)
- Copilot
- Other models

Coding Tools

- Cursor
- Claude CLI/Desktop
- Gemini code
- VS Code
- Codes

Note: some tools can be used with different models



What you need (Continued)

Python

- Most commonly used model
- The uv package manager is essential

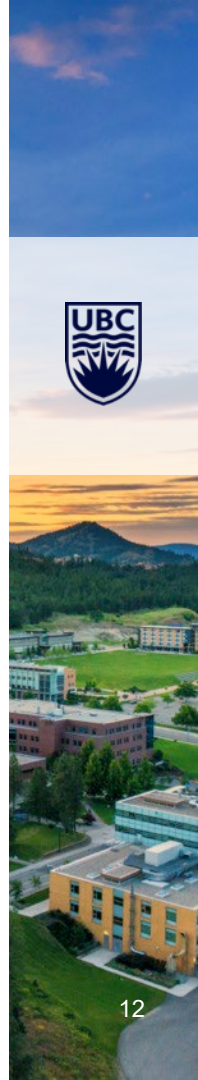
Git

- **Keep track of your AI interactions**

Willingness to explore

- Tools are showing up constantly:
  - https://github.com/ruvnet/claude-flow
  - Model Context Protocol (MCP) Explained - by Nir Diamant

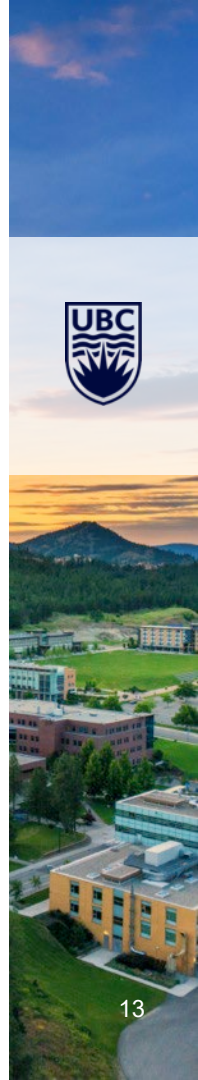
TAs can assist



What you need to do

Note: there will **not** be any assessment for this.

- Set up a local (under your control) AI environment
- Test our queries against your selected environment
- Document the interactions (prompt, response, your evaluation
    - Share these on Discord
    - Bring descriptions to class – this **will** be discussed
- What worked? What didn't work?



Building Systems that Track, Scale, and Manipulate

Background:

- *Las Vegas Skydiving v Groupon*
- I served as an expert for Las Vegas Skydiving
- Case involved using LV Skydiving's service mark ("Fyrosity") on Groupon's pages to create confusion.

Note: Las Vegas Skydiving **lost** this case, despite clear (to me) evidence that Groupon was in fact using their service mark (I don't think their lawyer used the information that I provided to them, based upon the record.)

Key Technologies

Groupon, like most web companies, seeks to constantly track where referrals come from.

- Groupon has local groups that actively seed links into non-Groupon customer sites (e.g., LV Skydiving's Facebook page)
- Groupon used *Bitly*

Bitly is a "link shortener" company

- Benefit: take a long link and convert it to a short link
- https://fsgeek.ca/teaching/cpsc-436c-2025-winter-term-1 -> http://bit.ly/3VEor8l
- Business model: analytics

The Lesson

Technical systems enable manipulation at scale

Corporate behaviour invisible to non-technical users

Understanding architecture means understanding power

Maybe a better lawyer would have won this case

- Suggests an experiment

Today's Mission

We are going to build something like what Groupon built

You will think it is "just a URL shortener"

You will learn to see how it is a surveillance system, based on cloud computing technology

Let’s Build It

Resources:

- AWS Free Tier
- Claude Flow
  - Let’s use locally hosted LLMs

Warning: the following slides will be done as **live demos** and while I’ve made sure they work for me, they might go wonky in class.

Demo

Let’s see what happens when we ask AI to do this for us.

Discussion

Too much background… Let's talk.


HighYield
The
Impossible Scale
of
AI Data
Centers


THE UNIVERSITY OF BRITISH COLUMBIA
THE UNIVERSITY OF BRITISH