CanvasGPT

Paris Phan, Ethan Zhang

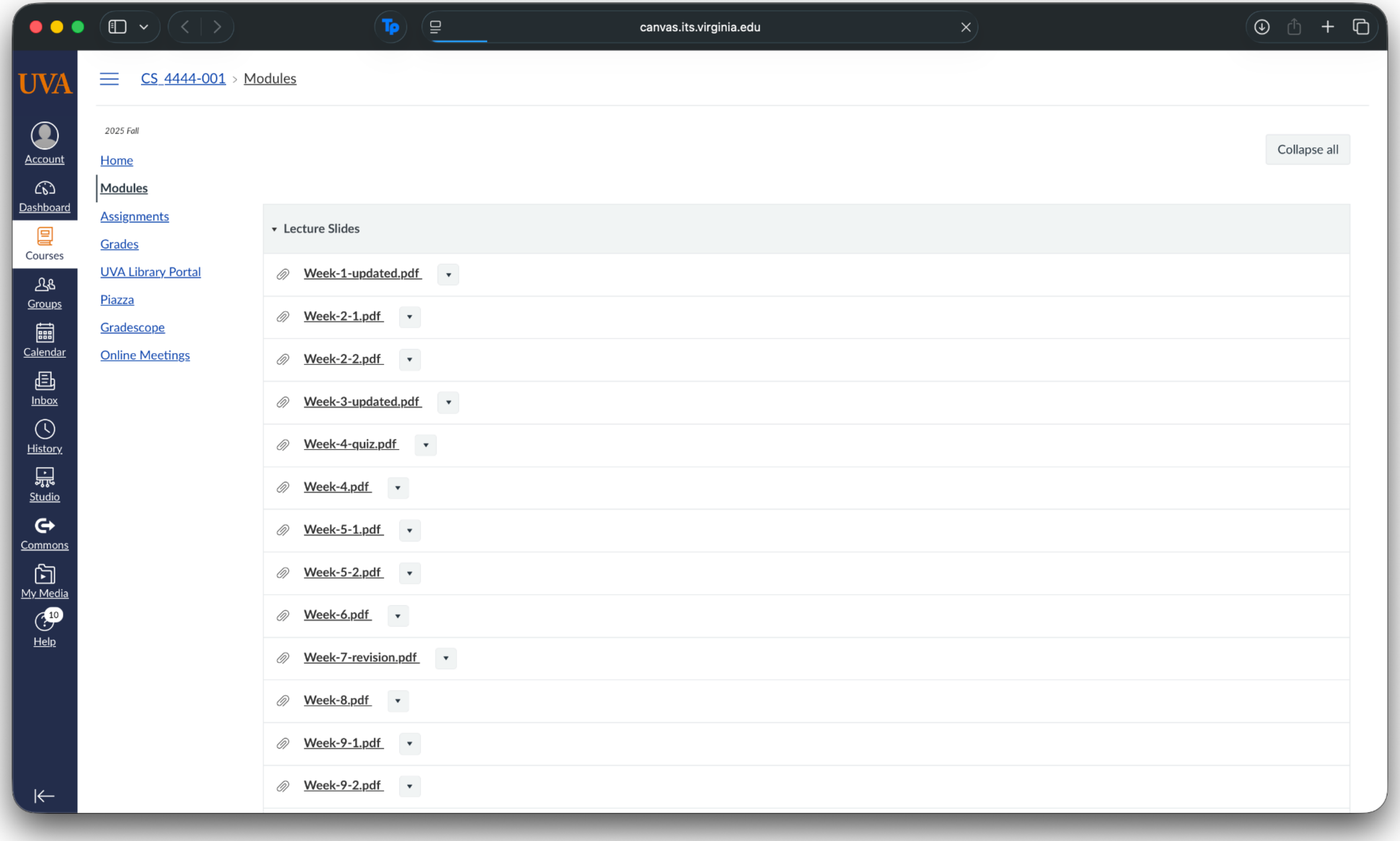
Nov 11, 2025

Background / Motivation

- Canvas offers educators a high degree of freedom in structuring their courses
  - Syllabus
  - Announcements
  - Grades
  - Gradescope

- Every teacher has a canvas layout that differs slightly
  - Some use announcements for assignment details instead of ‘assignment’ tab
  - Some separate assignment submissions from details, incorporating Gradescope
  - Some teachers use their own course website (e.g. github page or Webassign) for all information, and canvas becomes just a proxy

- Students have trouble keeping track of course structure and materials across 5+ classes in addition to possible external sites
- It can be difficult to ensure you’re on top of everything if you’re unable to find specific information



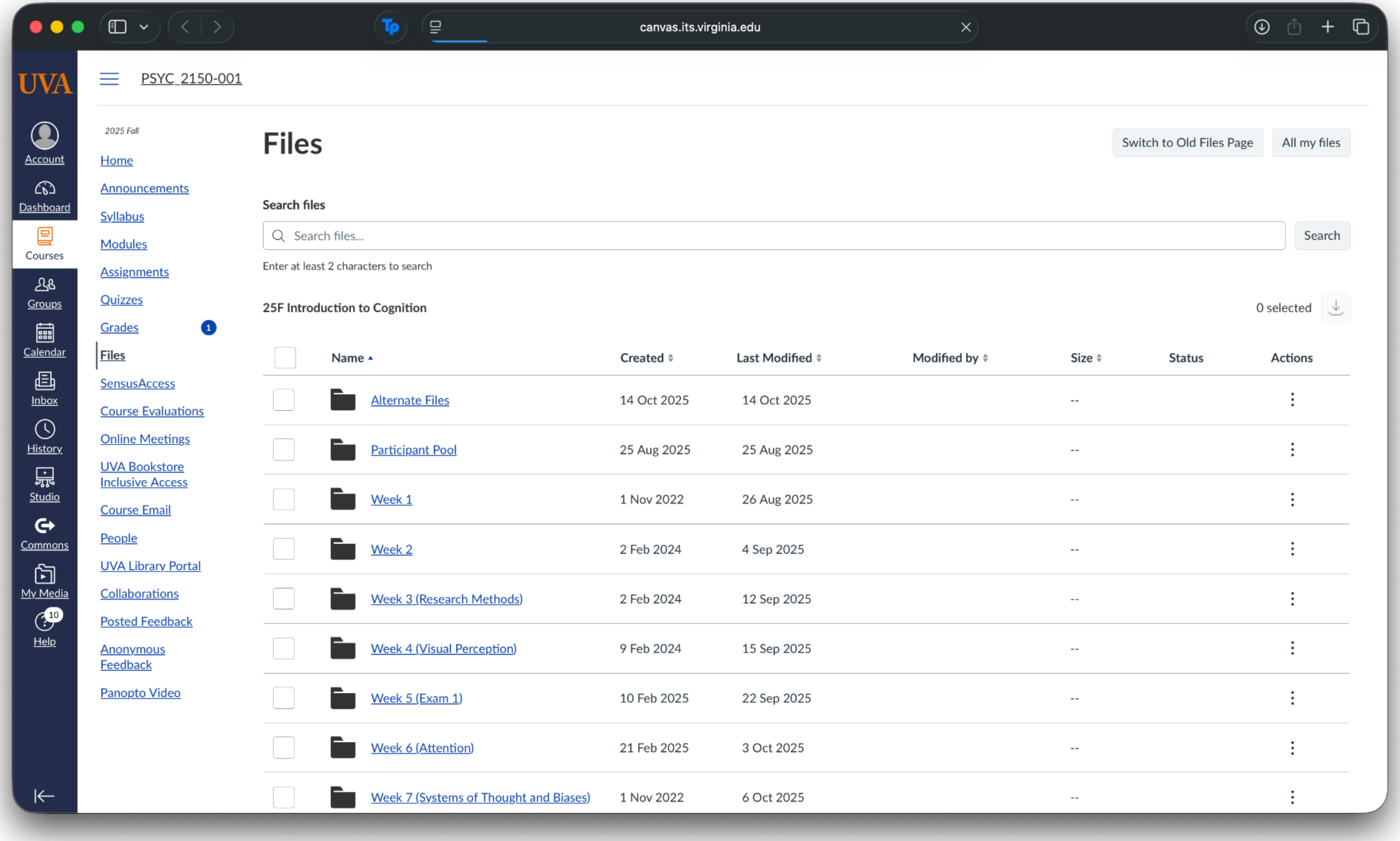
CS 4444-001 > Modules

2025 Fall

- Home
- Modules
- Assignments
- Grades
- UVA Library Portal
- Piazza
- Gradescope
- Online Meetings

▾ Lecture Slides

- Week-1-updated.pdf
- Week-2-1.pdf
- Week-2-2.pdf
- Week-3-updated.pdf
- Week-4-quiz.pdf
- Week-4.pdf
- Week-5-1.pdf
- Week-5-2.pdf
- Week-6.pdf
- Week-7-revision.pdf
- Week-8.pdf
- Week-9-1.pdf
- Week-9-2.pdf



Files

Search files

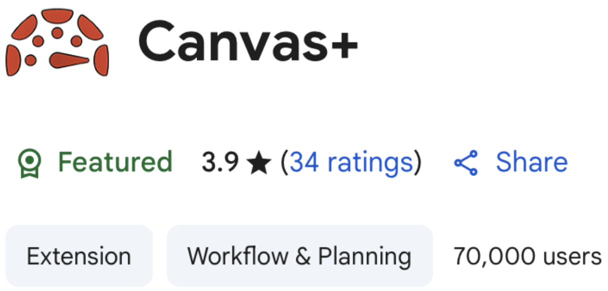
Search files...

Enter at least 2 characters to search

25F Introduction to Cognition

□	Name ▲	Created	Last Modified	Modified by	Size	Status	Actions
□	Alternate Files	14 Oct 2025	14 Oct 2025		--		⋮
□	Participant Pool	25 Aug 2025	25 Aug 2025		--		⋮
□	Week 1	1 Nov 2022	26 Aug 2025		--		⋮
□	Week 2	2 Feb 2024	4 Sep 2025		--		⋮
□	Week 3 (Research Methods)	2 Feb 2024	12 Sep 2025		--		⋮
□	Week 4 (Visual Perception)	9 Feb 2024	15 Sep 2025		--		⋮
□	Week 5 (Exam 1)	10 Feb 2025	22 Sep 2025		--		⋮
□	Week 6 (Attention)	21 Feb 2025	3 Oct 2025		--		⋮
□	Week 7 (Systems of Thought and Biases)	1 Nov 2022	6 Oct 2025		--		⋮

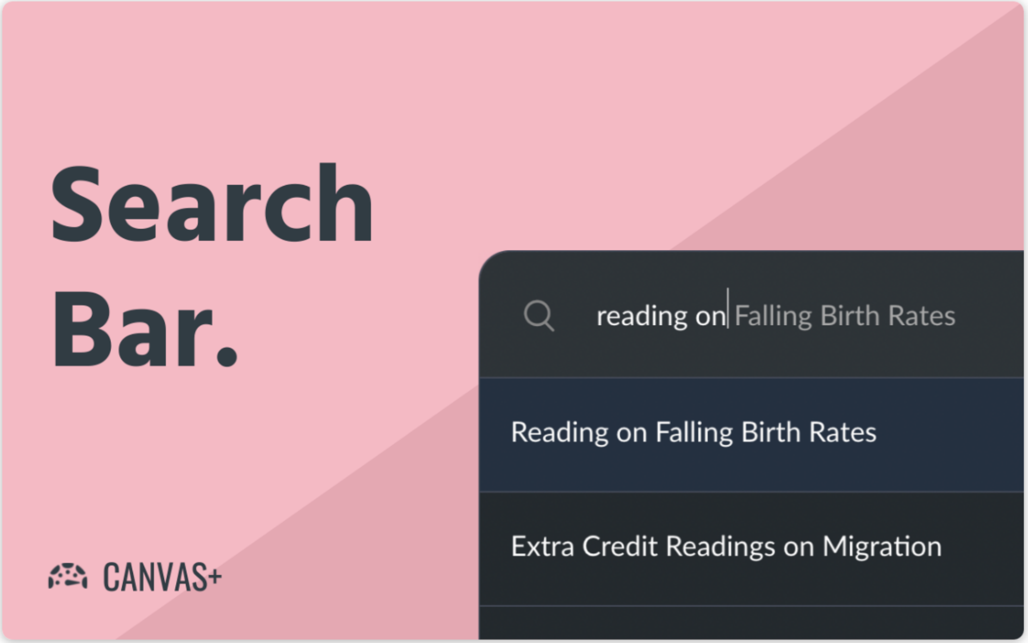
Related Work



Canvas+

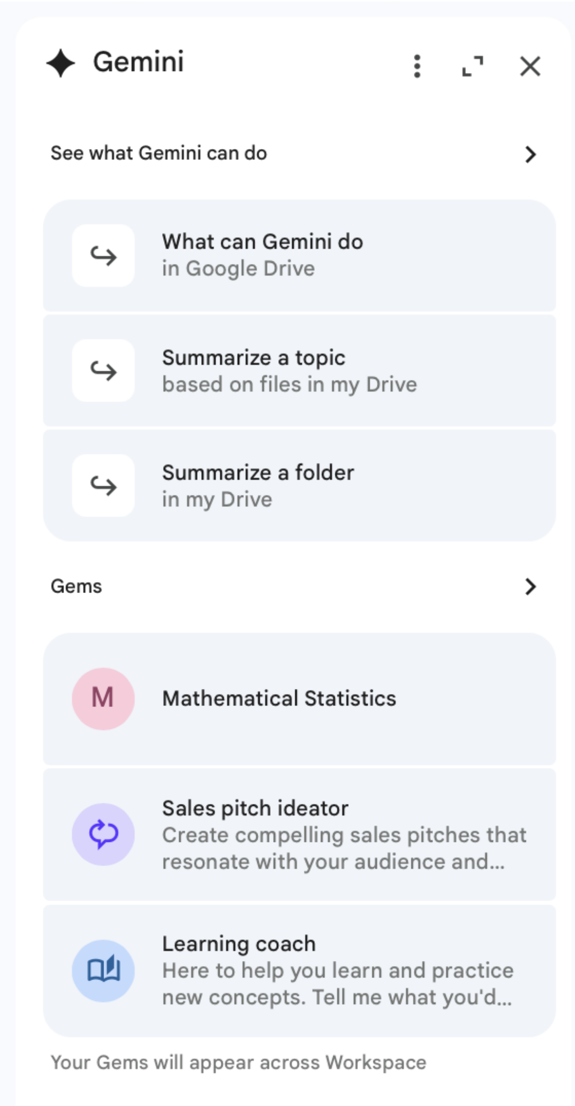
Featured 3.9 ★ (34 ratings) Share

Extension Workflow & Planning 70,000 users


Search
Bar.
reading on Falling Birth Rates
Reading on Falling Birth Rates
Extra Credit Readings on Migration
CANVAS+

- Offers slight convenience features
- No AI integration
- No Smart Assistance
- No Information compilation

Related Work



✦ Gemini

See what Gemini can do

What can Gemini do in Google Drive

Summarize a topic based on files in my Drive

Summarize a folder in my Drive

Gems

Mathematical Statistics

Sales pitch ideator
Create compelling sales pitches that resonate with your audience and...

Learning coach
Here to help you learn and practice new concepts. Tell me what you'd...

Your Gems will appear across Workspace

Massive amount of precedence in work platform AI indexing

- Many workspace platforms integrate AI indexing and search
- Google Drive
- Notion Workspace AI
- Slack AI
- Microsoft Copilot

Even local document-based assistants

- Warp (AI Terminal)
- Claude Code
- Cursor
- etc.

Claim / Target Task

To make life easier for college students with an interface for Canvas that unifies, normalizes, and semantically indexes course content and assignments across inconsistent instructor setups, enabling intelligent retrieval and proactive alerts for students.

Proposed Solution

An app / platform that connects to your Canvas student account, that utilises AI/ML to

- Read and understand all your courses and their individual unique structures
- Ingest and unify data to provide a single structure to view all your course information
- Automatically detects new updates to courses, including new assignments/grades/etc.
- Notifies you of new updates, upcoming deadlines, etc.
- (backlog) Integrates with your workspaces (Google drive, notion, etc.)
    - Automatically updates and fills your workspace with new material

Implementation

Electron Desktop App

- Utilising local ML models + API LLMs

3 Major Components

- LangGraph Discovery Agent
- Unstructured Data Discriminator
- MCP LLM Interface

LangGraph Discovery Agent

Agentic workflow to automatically discover

the routes of data sources for future data ingestion

- Starts at the canvas home page

- Determines valuable native canvas endpoints

- Explores linked external sites for missing sources

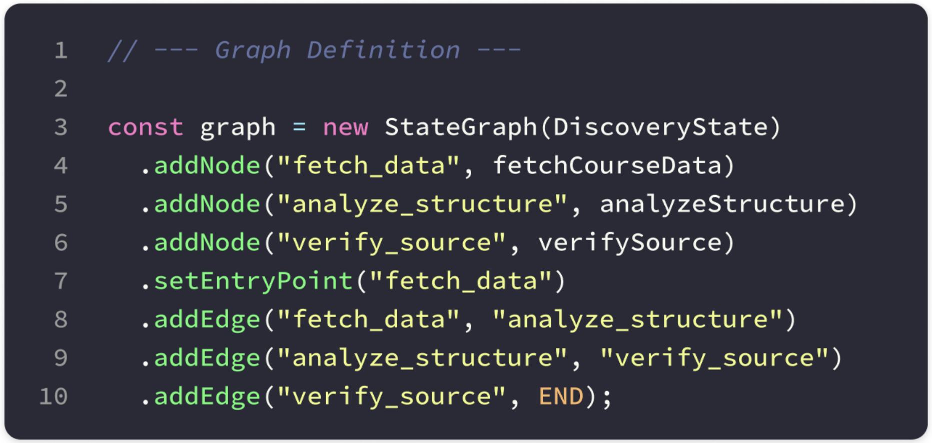
<u>Needed Data Sources</u>

Syllabus

Assignments

Course Learning Material

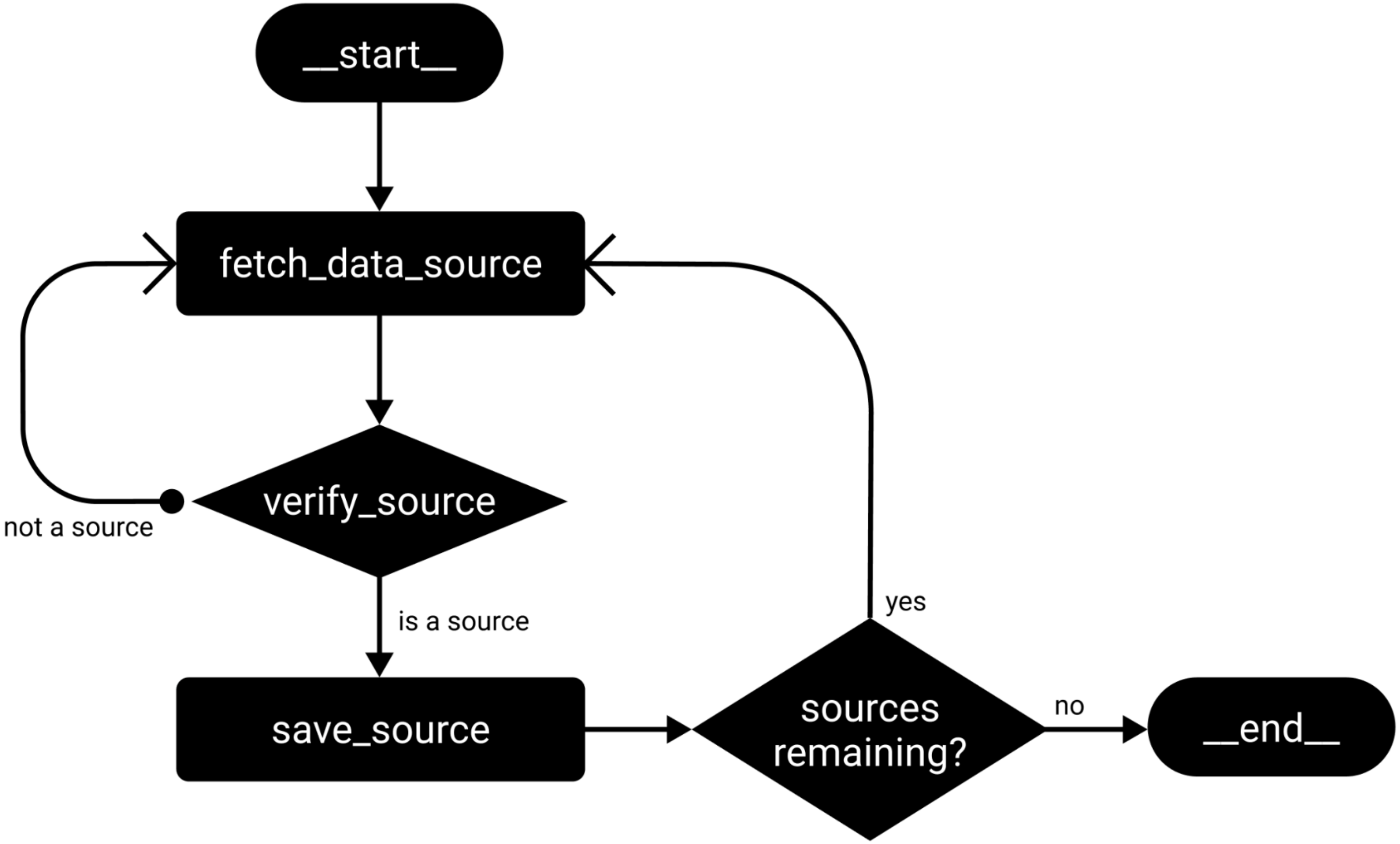
Schedule



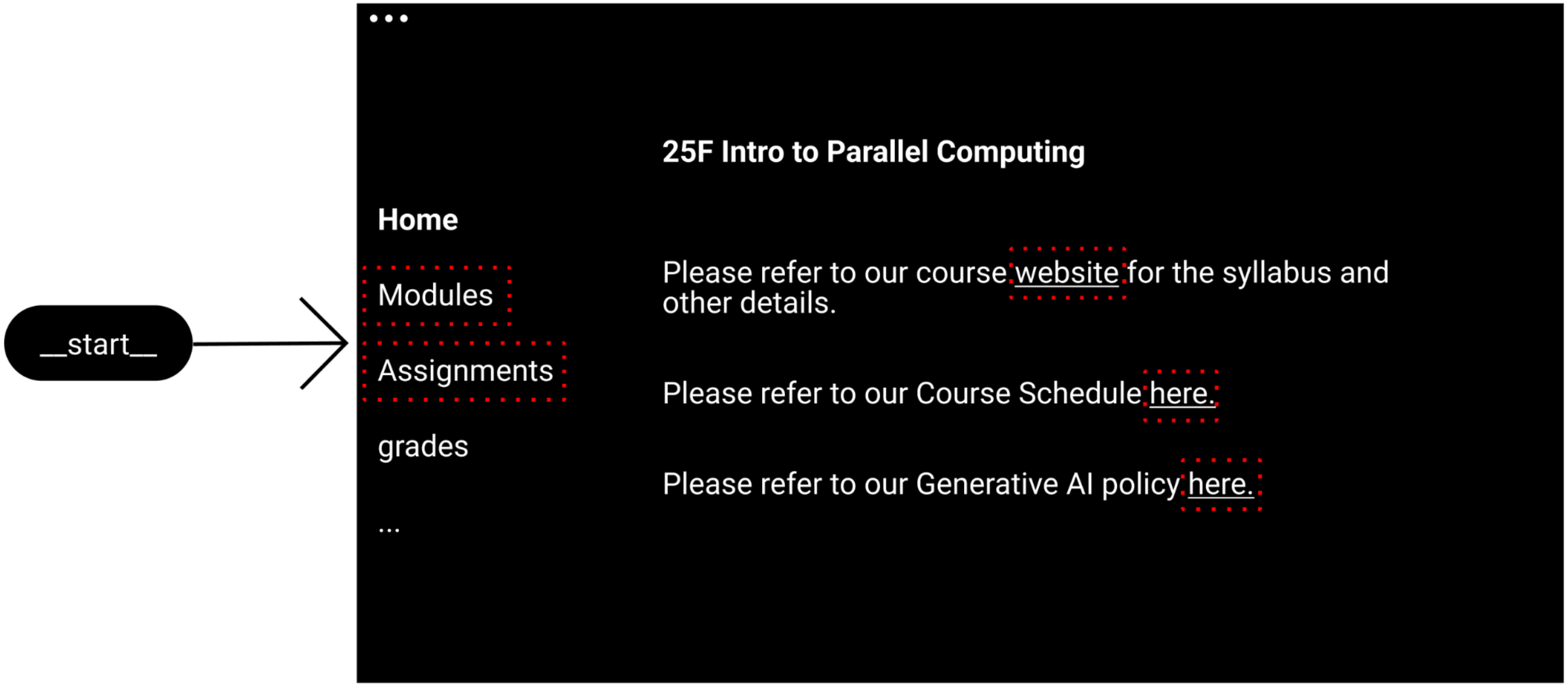
```
// --- Graph Definition ---

const graph = new StateGraph(DiscoveryState)
  .addNode("fetch_data", fetchCourseData)
  .addNode("analyze_structure", analyzeStructure)
  .addNode("verify_source", verifySource)
  .setEntryPoint("fetch_data")
  .addEdge("fetch_data", "analyze_structure")
  .addEdge("analyze_structure", "verify_source")
  .addEdge("verify_source", END);
```

LangGraph Discovery Agent



LangGraph Discovery Agent



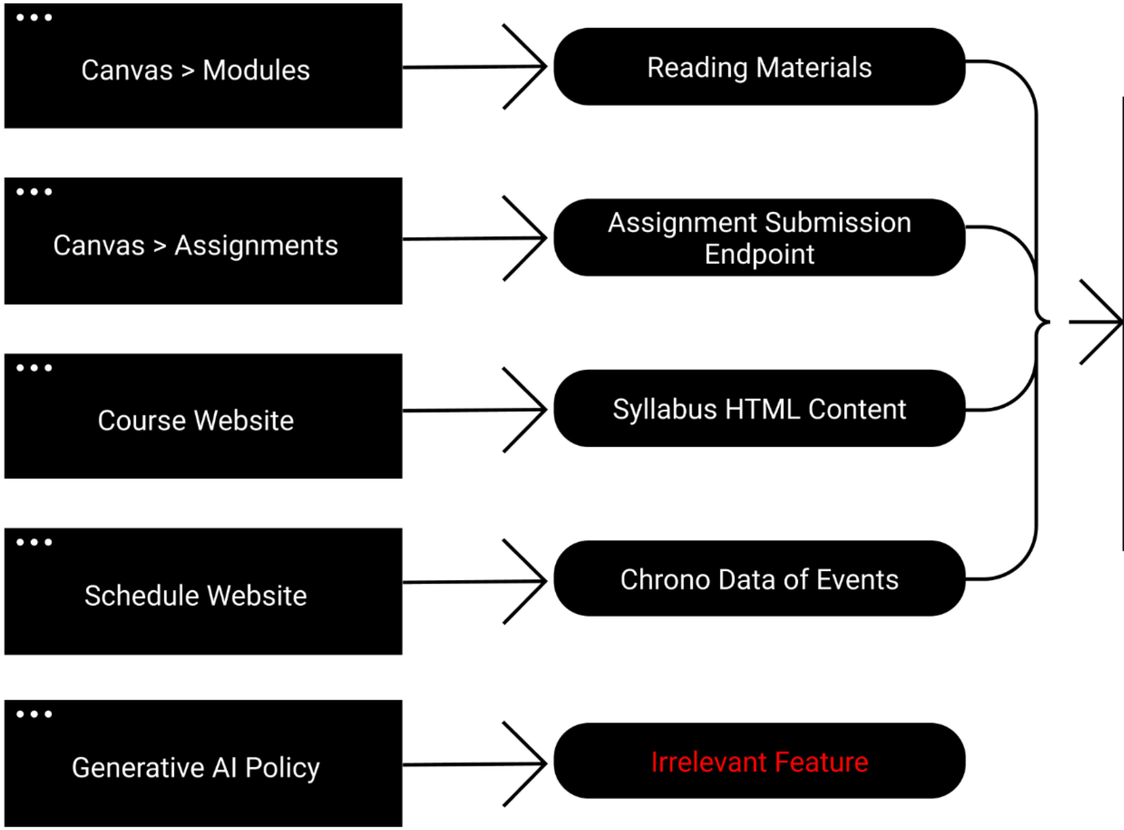
25F Intro to Parallel Computing

Please refer to our course website for the syllabus and other details.

Please refer to our Course Schedule here.

Please refer to our Generative AI policy here.

LangGraph Discovery Agent



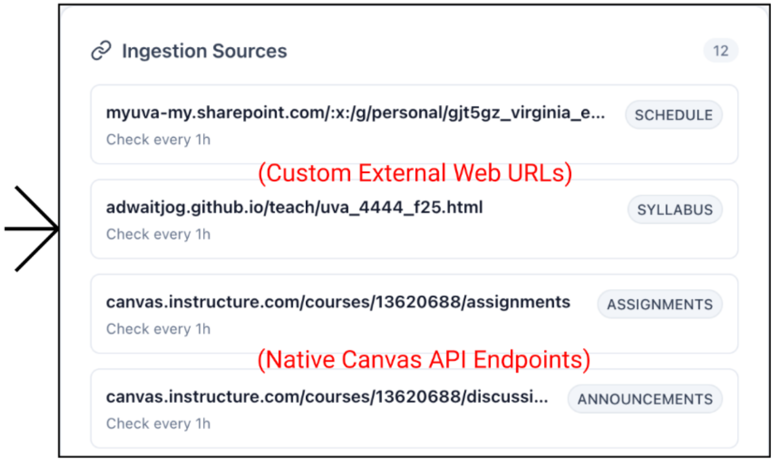
Canvas > Modules → Reading Materials

Canvas > Assignments → Assignment Submission Endpoint

Course Website → Syllabus HTML Content

Schedule Website → Chrono Data of Events

Generative AI Policy → Irrelevant Feature



Ingestion Sources 12

myuva-my.sharepoint.com/:x:/g/personal/gjt5gz_virginia_e... SCHEDULE
Check every 1h

(Custom External Web URLs)

adwaitjog.github.io/teach/uva_4444_f25.html SYLLABUS
Check every 1h

canvas.instructure.com/courses/13620688/assignments ASSIGNMENTS
Check every 1h

(Native Canvas API Endpoints)

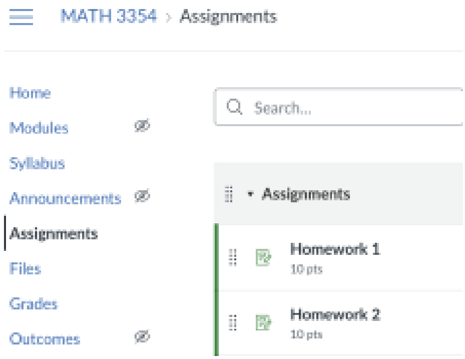
canvas.instructure.com/courses/13620688/discussi... ANNOUNCEMENTS
Check every 1h

Data From Ingestion Sources

We have all possible data for a course

BUT, they are all unorganised and unstructured

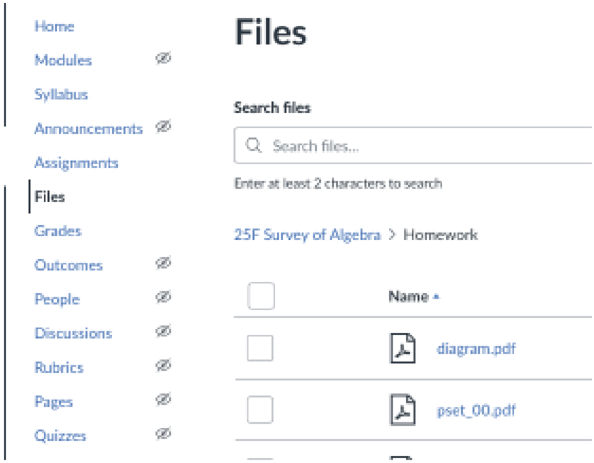
For example:



Canvas > Assignments

Gives us the submission endpoint for our assignments, but **<u>no due date</u>**, **<u>no task info</u>**

Instead, this information is found in a pdf, in Canvas > Files!



Unstructured Data Discriminator

Each piece of data is stored as a ‘universal item’ in our database

- Highly unstructured, no item rules
- [type: Assignment, name: Homework 1, metadata: NULL]
- [type: File, name: pset-1.pdf, metadata: {due_date, desc, ...}]

We need to find a way to connect these two pieces of data and form one true **realised entity** to show the user.

- Confident in its existence
- Confident in its metadata (correct due date)

Unstructured Data Discriminator

Phase 1: Feature Extraction & Vector Embedding (?)

- extract ‘assignment number’ from every string using regex

Phase 2: Clustering Algorithm

- group all related items together (e.g. pset01.pdf, homework 1)

Phase 3: Verify confidence levels

- Prompt user to ‘triage’ items with confidence < 0.9

MCP LLM Interface

3 Major Components

- MCP LLM Interface
- Host/Expose the Electron App as a MCP Server
- Claude Code is able to access this MCP server, calling the exposed tools
- Gives Claude the context of all the canvas ingestion
  - Can help with Homework
  - Plan out homework
  - etc.
- Will extend to other tools (Cursor, Notion, Zapier, IFTT, etc.)

Challenges

Unstructured Data Discrimination Model

- Rule Based Fuzzy Matching Algorithm
    - Lacks semantic understanding for unstructured courses
- Discriminative Entity Recognition (Random Forest, XGBoost))
    - Requires training dataset
- Feature Extraction and Clustering
    - Timeline Feasible

Future Improvements

Unstructured Data Discrimination Model

Hybrid Graph-Based Resolution Approach

1. Feature Extraction & Blocking
   - Rule matching & Normalisation
2. Candidate Generation
   - SBERT Semantic Embeddings to compute cosine similarities
3. Transitive Graph Resolution
   - Union-Find to find connected components

Application Package

- Electron App (Electron-Vite)
- LangGraph
- openAI
- MCP
- Claude Desktop
- LanceDB
- TailwindCSS

Any Questions?