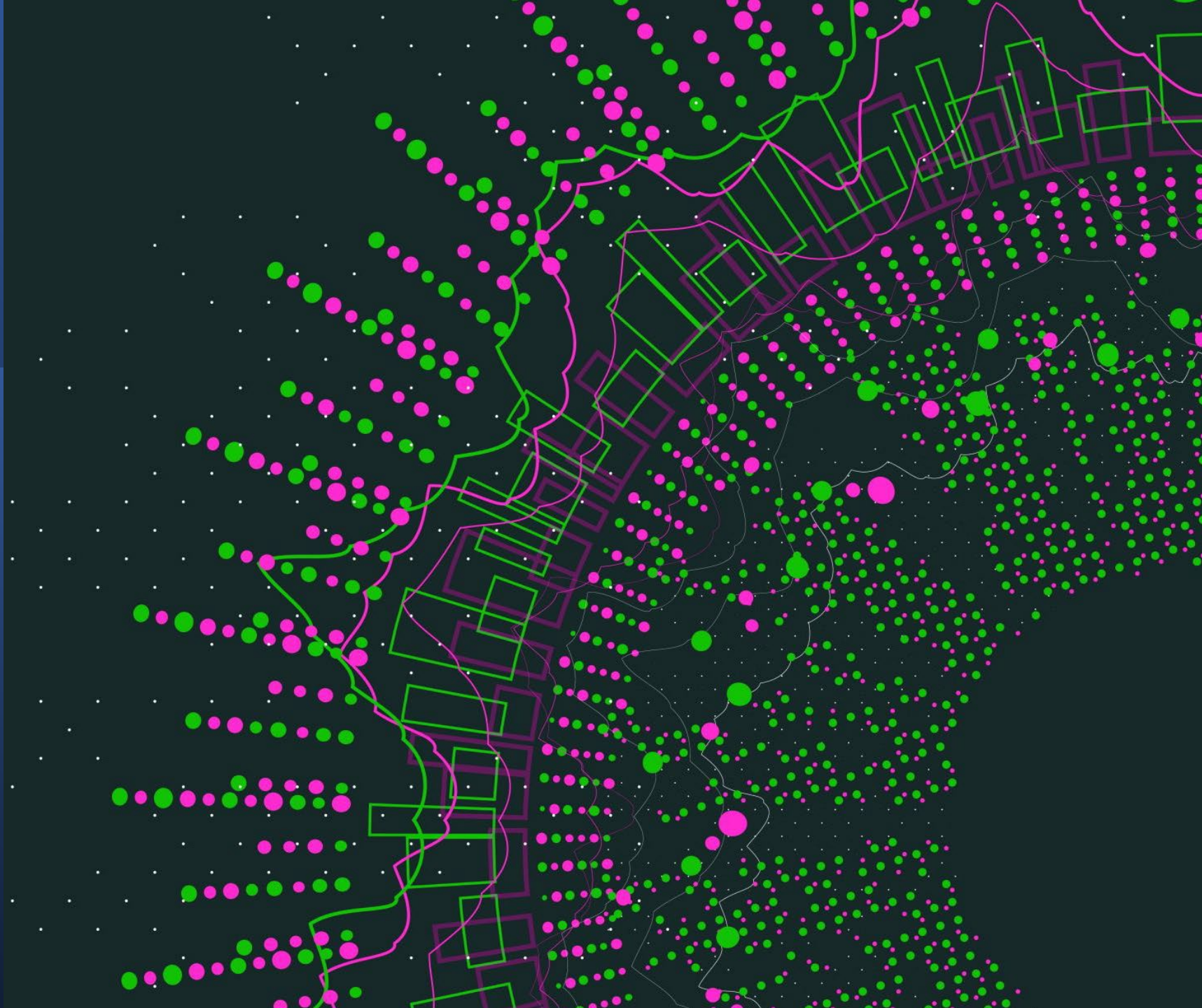Rajiv Garg

Summer 2023 Research

RA Assignments

- Paid positions (only for currently enrolled students at Emory): 10hr/week for the entire summer (about 12 weeks)
  - You will need to submit a status report every week for the tasks completed during the past week
    - keep track on Google Docs but share the new updates by email
  - You must submit a timesheet with hours worked every two weeks.
    - Please keep the hours worked during any week as 10hrs/week for simplicity and consistency (you will spend less than 10 hours some weeks and more than 10 hours other weeks on research)
    - Avoid filling in work hours for lunch (12-1) or late at night (after 12 am).
      - Requires additional approval delays.

- Unpaid positions: 20hrs/week for a negotiated time period (FLEX-TIME AND TASK-DRIVEN)
  - You will need to submit a status report every week for the tasks completed during the past week
    - keep track on Google Docs but share the new updates by email

- Slack/Discord?
  - https://discord.gg/KnkM9MWU
  - https://join.slack.com/t/aigarg/shared_invite/zt-1v14zmbsz-Z083cJ7UcRBNtpN37p015g
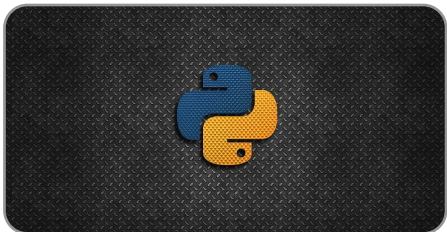
Summer Projects

Generative AI Research Assistance

- AdAI: Social media text ad (generative AI), effectiveness, and bias (currently: Jessica)
- VideoAI: Video ad generation (generative AI text script, synthetic voice, and image) effectiveness, and bias (currently Katie)
- VoiceAI: generate voices/videos with provided scripts and execute Mturk (market research) study
- EdAI: generative AI lecture vs human lecture (measure student performance)
- ProductAI: Production description engineering for channel placement

Text Analytics and Data Science Research Assistance

- AdMatching: personality-based ad matching for effectiveness
- FakeIt: text analysis of videos used in a fake content experiment
- EdAdopt: role of tech adoption intensity in K-12 education on student performance
- TechSkils: feature extraction from user resume for prediction of job loss
- DeepStar: Image feature extraction and detection for star-rating
- CryptoTax policies

Business Analytics Research Assistance

- Book review (*PAID*)
- Paper review
- Paper writing
- Big data course website and resource compilation (github)
  - Auto-grader development and test
  - E-book creation

Generative AI projects (mid/late stage)

- **AdAI**: Social media text ad (generative AI), effectiveness, and bias (current RA: Jessica)
  - Create and deploy generative AI text ads to target specific user demographics.
  - Execute on Facebook/Twitter/Google/etc.
  - Need someone with <u>experience with running advertising campaigns</u> to work with Jessica

  use chatGPT to create ads; have the code, currently in testing state (for different demographics); just dealing with companies and setting up the experiments.

- **VideoAI**: Video ad generation (generative AI text script, synthetic voice, and image) effectiveness, and bias (current RA: Katie)
  - Create and deploy video ads to target specific user demographics
  - Execute on YouTube and Instagram/Facebook/TikTok.
  - Need someone with <u>experience with video creation and willing to learn deployment of advertising campaigns</u> to work with Katie

  chatGPT writes the text, other api vocalize the ads, then put together a video; see if it is working, no data, need to create videos and run campaigns on different platforms.

- **VoiceAI**: generate voices/videos with provided scripts and execute a Mturk (market research) study
  - Create and deploy videos on Mturk to execute A/B testing
  - All details are already in place, and a pilot study was already executed – you will need to learn the system and run a larger-scale study

  create different kinds of voices, clustering from 60,000 voices. 6 different kind of voices, same script, need to create videos, run on youtube, and put on mturk, AB testing.

Generative AI projects (early stages)

chatGPT writes slides and compare it to online
courses. need to create courses and deploy it.

- **EdAI**: generative AI course vs human course (measure student performance)
  - Identify TWO popular online courses – one technical (e.g., SQL?) and one non-technical (marketing strategy?)
  - Generate four sets of video lectures (on some online learning platform or YouTube) – one using existing slides (E) and one using content created by ChatGPT (C), one using AI voice (A) and one using human voice (H).
  - Do A/B testing to estimate the effectiveness (engagement/views/likes) of all four combinations: E+A, E+H, C+A, C+H

- **ProductAI**: Production description engineering for channel placement
  - I have already classified product descriptions that are good for online/offline channel placements
  - Extract features from these descriptions
  - Use the combinations of features to develop a create (using chatGPT) product description based on the expectation of product placement
  - Test the product descriptions using a lab study or MTurk

determine if better online/offline (already finished). If want to sell in doors, how should they write the
descriptions? Use chatGPT, give features, let it generates a good product description (for specific
channel)???
Do you need to touch/feel the products to buy?

Text Analytics and Data Science (mid/late stage)

- **AdMatching**: personality-based ad matching for effectiveness
  - Using provided Tweet data to estimate brand personality (using provided Python code)
  - Identify and trim 30-second video clips and attach banner ads to the videos
  - Deploy those videos on YouTube and execute a lab/Mturk study to test the performance of ads

- **FakeIt**: text analysis of videos used in a fake content experiment
  - Using provided YouTube videos, extract the script using OpenAI Whisper or other speech-to-text APIs
  - Using Python, extract topics from those video scripts
  - Perform empirical analysis using the data

  extract text from videos, see if
  fake data is more effective
  extract topics, see if time series matter

- **CryptoTax**: Crypto tax policies (current RA: Qing)
  - All data collected. Merge data and check if any data is missing.
  - Create visualizations using Tableau or Python – there is time-series data and text data.
  - Run regression models
  - Document the findings

  tax policy's effect on crypto price
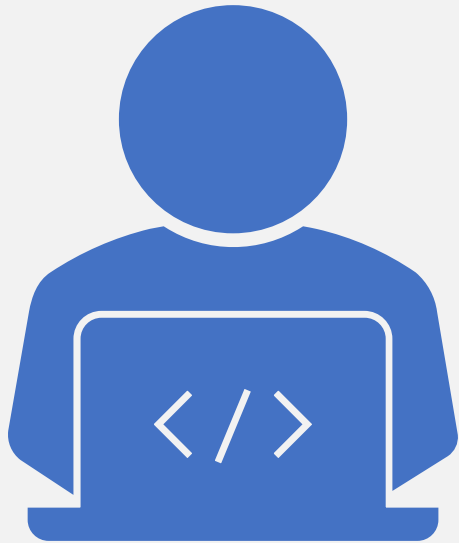 already have the data, need insights

Text Analytics and Data Science (early stage)

- **EdTechAdoption**: role of tech adoption intensity in K-12 education on student performance
  - Already collected data from a few thousand schools tech adoption during pandemic
  - Clean the data and identify all tech adoption during covid lockdown
  - Download school performance data and run regression models

- **TechSkills**: feature extraction from user resume for prediction of job loss
  - Already collected data on a few hundred employees using webscraper.io (past RA: Vijay)
  - Need to collect additional data
  - Perform feature extraction from user profiles
  - Develop a classification model for job loss prediction

- **DeepStar**: Image feature extraction and detection for star-rating
  - Start with Yelp image data. Extract features from images and predict star ratings based on features.
  - If reasonable accuracy, we will get product images and re-perform the tasks to predict star ratings
  - Extension to Instagram influencer images

Writing and Reading Projects

- Book review (*paid position for all*)

- Paper review for stylesheet (DUE: MONDAY)

- Research paper writing (various topics)

- Big data course website and resource compilation (GitHub)
  - Possibly create a short e-book for the course (you can use chatGPT)
  - Auto-grader development and test
    - https://gradescope-autograders.readthedocs.io/en/latest/mysql/

QUESTIONS?

- Email me the following information (if you have to leave today):
  - Emory student status during summer 2023 (currently enrolled or not)

  - Top TWO projects you are interested in

  - Weekly group meeting (1hr) day preference – Thursday (at Noon) or Friday (at Noon)

  - Interested in teaming up with someone?

  - Do you need a formal offer letter?

  - Available next week on Friday for an interview if not available today?