Second Project: Embodying Algorithms in Software
Law and Algorithms, Spring 2024

A. BU's AI Task Force

Generative AI has captured the imagination of just about every field, including higher education. Technically speaking, these tools are simply machine learning algorithms that are capable of creating output that convincingly imitates the products of human effort, by using probability to guess which words or other content should follow from a given prompt given their gargantuan training data. But the degree to which they can produce convincing results has many universities rethinking many of the fundamentals around how we approach university education and administration.

On January 29, 2024, BU President Ad Interim Kenneth Freeman sent an email to all members of Boston University where he announced the creation of an "AI Administrative Process Task Force" to explore the appropriate and inappropriate uses of AI software by faculty, staff, and administrators in their official duties. The task force is assigned to "explore, evaluate, and propose AI solutions in various administrative functions across the University" in order to "create a more agile and responsive administrative framework while managing related risks."

Note that this task force is *not* tasked with creating policies that govern the use of generative AI by students when completing assignments; many colleges and departments in BU already have such policies in place. We reproduce the full email contents in the Appendix below.

The task force is required to present their findings back to the university's AI Task Force by June 30, 2024.

All of the above is true. Here's where the fictional part of the assignment begins:

B. Your Assignment

The co-chairs of the AI Administrative Processes Task Force, Bob Graham (Assistant Vice President, ERP, CRM & Integration, IS&T) and Kelly Lockard (Assistant Vice President, Continuous Improvement & Data Analytics), have enlisted your group to provide assistance on a specific matter: how the BU administration can detect whether a given piece of text is the product of a large language model or was written by a human. This is useful for detecting AI-generated content in student work, but BU would like to develop a system that can scan and detect this in a variety of media generated around and on behalf of the

university. That way, if BU faculty or staff are detected using AI generation tools in situations where such use is inappropriate, their supervisors can be informed and can take corrective action.

To help with this, Boston University is conducting negotiations with a startup company called d3tect.ai.[1] D3tect.ai's detection technology can be set up to automatically scan content within BU communications, marketing, and publication platforms with minimal interruption to how those platforms operate today—including how content is posted to bu.edu, on BU-controlled social media, in press releases, and as is posted by faculty on platforms like Blackboard and Piazza.

If a piece of content is flagged as having been generated in substantial part through the use of a machine learning algorithm, an email will be automatically sent to the author's supervising staff member or Dean. That supervisor will then set up a meeting to review the incident and take appropriate corrective and disciplinary responses. D3tect.ai plans to have a deployable version of its platform ready by Fall 2024.

News about this Task Force's consideration of d3tect.ai led to significant concern and pushback among various sectors of BU faculty and administration, including the BU Faculty Council, the local chapters of the UAW and SEIU representing various BU staff, and the BU chapter of the American Association of University Professors. In particular, there were concerns that the technology would incorrectly flag faculty and staff content, resulting in unwarranted investigation and discipline.

In response to these complaints, the Provost's Office has allocated a budget of $250,000 to invest in processes and technologies that will help faculty and staff feel more comfortable with the deployment of d3tect.ai. They are ready to purchase these tools and services immediately, giving about a seven month period of assessment before d3tect.ai's tool is actually deployed in September 2024. However, the Task Force has made clear that deployment of d3tect.ai's technology will start in Fall 2024—with or without any additional trustworthiness mechanisms.

C. The Budget

You are being asked to submit a recommendation for how the administration should spend its $250,000 budget to allay student and faculty concerns.

---

[1] There are several real companies and tools that are developing this technology that you could use to learn about how this fictional company might be approaching the problem. For example, see https://originality.ai/, https://gptzero.me/, https://writer.com/ai-content-detector/, http://gltr.io/, https://quillbot.com/plagiarism-checker, and turnitin.

Boston University has negotiated the following prices with d3tect.ai for a number of different technical options that they could arrange:

- ***Verification and validation (choose from one or more below).*** Before rolling out d3tect.ai's technology, members of the BU community are able to select a set of "test vectors." Specifically, members of the BU community (selected using the criteria below) can create a set of specimen content for the algorithm to screen. These will be varied by "origin," either human generated content or content generated with significant help from a machine learning model, and the "label," or type of content (e.g., an essay, a social media post, photos, graphics, etc.). Each time d3tect.ai updates its technology, they will publicly release the results of their classifier on these test vectors. Once d3tect.ai's tech is deployed, however, no additional test vectors can be added.
    - ***(Option 1: $40,000)*** The University will hire 4 graduate students at 20 hours per week for a semester to generate the test vectors. If you select this option, please determine how these students should be selected.
    - ***(Option 2: $60,000)*** The University will pay a professor in a computational field (e.g., CDS, CS, Statistics, Math, Economics, etc.) to generate the test vectors, replacing their teaching and service responsibilities for a semester. If you select this option, please determine how the faculty member should be selected.
    - ***(Option 3: $150,000)*** The University will pay an external auditor from local well-reputed, Boston company BeantownAuditors to generate the test vectors.

- ***"Deposition" of d3tect.ai's Chief Technology Officer ($45,000)***. d3tect.ai has agreed to allow a team from BU's Faculty Council and Local 888 of SEIU (which represents a variety of clerical and technical workers around BU) to take an 8-hour voluntary interview (conducted like a deposition) of its Chief Technology Officer, Patty Whealan, who oversaw the team of three engineers who created this tool. She would answer questions exactly as one would in a deposition, and swear an affidavit at the conclusion that, to the best of her knowledge and under the penalty of perjury, all answers provided were true and not misleading.

  Additionally, she has agreed to produce the company's design documentation and internal Slack communications to the deposition team under a protective order, but there would be no access to d3tect.ai's source code.

  The investigatory team would then be allowed to release some or all of the

transcript to the staff, students, and faculty at BU, who would be permitted to reference it in both public debates and in any disciplinary proceeding. Per the protective order, they would not be able to release copies of any of the underlying documents referenced during the interview.

- ***Confidential adversarial access (select one only below).*** Boston University will amend the current disciplinary procedures for staff and faculty such that those accused (or their representatives, discussed below) are able to get access to d3tect.ai's source code (with comments), design documentation, and internal Slack communications under a protective order.

  Specifically, prior to any hearing or proceeding related to staff or faculty discipline, the accused and their representatives can sign a protective order and access (1) a Github repository with the source code and documentation of the currently deployed version of d3tect.ai system, and (2) a scanning API end-point for the deployed version of the system (i.e., the same access point that is used to scan BU content).

  Per the protective order, they will not be able to share any information learned through this access outside of the disciplinary proceeding or any subsequent litigation. The cost of this infrastructure depends on the level of access that BU representatives have to the platform:
  - ***(Option 1: $40,000)*** If this option is selected, only the accused themselves will get access to the documents above and scanning API end-point. This is a one-time cost, no matter how many staff or faculty require this level of access.
  - ***(Option 2: $80,000)*** If this option is selected, both the accused and a chosen representative will get access to the repository and scanning API end-point. The representative could be another member of the BU community, or an external expert. Again, this is a one-time cost, no matter how many cases are reviewed.

- ***BU-Specific Always-on API (select one only below).*** Everyone with a BU login will have query access to d3tect.ai's scanning technology through a special "test" portal, based on the specifications in the options below. Each user can make any number of queries to the scanner, but must wait at least 5 minutes between each query. Additionally, there is a maximum cap of 10 million test queries made per academic year.

  No record is kept of what queries were made and what responses were received,

and the portal is the same software and hardware as the version deployed for actual review for use of generative AI.

- ***(Option 1: $100,000)*** If this option is selected, the portal will provide the standard "likely generative AI use" / "no generative AI use found" response that it does when in operation.
- ***(Option 2: $125,000)*** If this option is selected, the portal will provide the standard response, as well as the internally-generated confidence value (expressed as a percentage) that it is correctly categorizing the output.
- ***(Option 3: $135,000)*** If this option is selected, the portal will be run with DEBUG mode turned on. This will provide (1) the response, (2) the confidence value (described in Option 2), and (3) a debugging output where any errors in how the software ran are noted and can be reviewed alongside the program outputs.

- ***BU-Wide Access Under Protective Order (choose from one or more below).*** Everyone affiliated with BU will have the option of signing a protective order in order to get access to pieces of d3tect.ai's internal information, as described below. Specifically, any member of the BU community can sign a statement at the Provost's Office, at which point they will be given credentials to access a repository with the particular documentation. The protective order prohibits any public disclosure of anything inside of the repository, but does allow staff and faculty to discuss any findings that they might have about the technology at any supervisor meeting, misconduct hearing, or subsequent litigation.
  - ***(Option 1: $100,000)*** The entire, original source code, with all comments and other internal documentation omitted.
  - ***(Option 2: $20,000)*** The in-code comments and other meta-information in the source code. (This will only be comprehensible if you also have access to the source code itself.)
  - ***(Option 3: $20,000)*** The binary code of all external libraries, dependencies, and other third-party software that is referenced in the d3tect.ai source code.
  - ***(Option 4: $30,000)*** The design documentation and software schematics that lay out the overall intended operation of the program.
  - ***(Option 5: $40,000)*** The internal d3tect.ai Slack logs for the channels used to develop this tool.

- ***Bug Bounty Program ($25,000).*** BU will pay the setup costs for d3tect.ai's bug reporter system. Specifically, d3tect.ai will make a bug reporter available, and encourage any person who finds an issue to report it to the company. In the event that any significant bugs are found through the bug reporter, those results will be

released to the full BU community after a 90-day embargo window (for d3tect.ai to review the findings and make any changes necessary). A third party auditor, BeantownAuditors, has agreed to review submissions in the bug reporter and classify whether or not a bug is "significant."

- ***External Company Audit ($250,000).*** Boston University will contract with Boston-based code auditing consulting company, BeantownAuditors, for a full audit. D3tect.ai will allow BeantownAuditors full access to their software, design documentation, developer Slack channel, and other business records. The company will do a 3 month review of d3tect.ai's existing technology and release a public report on their findings, over which d3tect.ai has agreed not to assert any right to review ahead of publication. Every time d3tect.ai releases a major version of their software over the next 5 years, BeantownAuditors will review the update and issue a report.

- ***Public Code Repository ($250,000).*** d3tect.ai is extraordinarily hesitant to make its source code available to the general public, but after extensive negotiation they have agreed to release their current model only with substantial payout. This will allow any member of the public to access the currently deployed model, including all parameters and hyperparameters. D3tect.ai will also make a bug reporter available under similar terms as the Bug Bounty Program above. In order to ensure that their business model remains viable, d3tect.ai will not publish other parts of their source code, including platform integration tools and the infrastructure they use to actually run the detection model.

D. The Assignment

Please write a report to the co-chairs of the AI Administrative Process Task Force indicating how you think the $250,000 budget should be spent, and justifying any choices that you make. Specifically:

1. Please list ways in which you believe the proposed integration of d3tect.ai's technology might fail to live up to its promise, the effect of these failures on the staff and faculty misconduct process, and the wider ramifications for the BU academic community. Try to be as comprehensive as you can.
2. Provide a suggestion for how to spend the allocated $250,000 to increase the trustworthiness of d3tect.ai's system. Explain why you think this particular set of transparency measures addresses the issues raised above (and to what extent it addresses these issues). Make sure to not only engage with each particular choice,

but the set of selections as a whole (e.g., how do the set of choices complement each other or offset the others' limitations?).

3. The University has another round of negotiations with d3tect.ai coming up on March 1, 2024. Are there any other transparency measures that are a valuable alternative for which BU should negotiate? Please explain and defend those alternative suggestions.

E. Logistics

We will be assigning groups of three to four. These will be randomly assigned and shared with the class.

Your assignment should be in the form of a report to be submitted to the BU AI Administrative Process Task Force addressing each of the questions above. Your submission should be between 2000–2500 words, citations excluded.

Please email your team's submission as a .PDF file to the course instructors (sellars@bu.edu and varia@bu.edu) before our class (i.e., before 2:10pm ET) on Feb 29.

F. Evaluation Criteria

Please review section 10 of the syllabus for our expectations for team collaboration for this and other assignments. Our grading rubric for this assignment will assess:

- How well the report explains the ways in which this software's deployment could fail to live up to its promises and the harms that could flow from unreliable technology in this area.

- Whether the team stayed within the $250,000 budget, and how it defended their selections in a way that engaged with software accountability debates.

- The quality of proposed alternative transparency measures, and how those alternatives are described and defended.

- The report's engagement with the assigned readings from Classes 4–6, and course discussion.

Appendix: President Freeman's Memo

Boston University Office of the President
Kenneth W. Freeman, President Ad Interim
One Silber Way
Boston, Massachusetts 02215

January 29, 2024

Dear Colleagues,

As we continue to witness rapid advancements in artificial intelligence, it becomes increasingly evident that incorporating AI into administrative processes is crucial for enhancing efficiency, effectiveness, and innovation. At the same time, the use of AI must be managed to avoid pitfalls such as inaccuracy and bias. With this in mind, the AI Administrative Processes Task Force is being established at Boston University.

The AI Administrative Processes Task Force will complement and collaborate with the Boston University AI Task Force, created last fall by the Office of the Provost, comprising faculty from across the University. That task force is focused on developing best practices and shared approaches around the use of generative AI in support of our research and teaching mission.

The primary objective of the AI Administrative Processes Task Force is to explore, evaluate, and propose AI solutions in various administrative functions across the University. This initiative aims to streamline processes, improve decision-making, and create a more agile and responsive administrative framework while managing related risks.

The AI Administrative Processes Task Force will be charged with conducting a thorough assessment of select administrative processes to identify areas that show the greatest potential to benefit from AI integration and to research and evaluate established and emerging AI technologies that align with the University's administrative needs. The task force is expected to recommend pilot programs to test the feasibility and effectiveness of AI solutions in select administrative areas and to develop training and support for staff members to ensure a smooth transition to AI-integrated processes and foster innovation in the use of AI. In addition, the task force will identify and recommend paths to elevate questions regarding the legal, ethical, and safe use of AI to ensure consistent and informed implementation, and to provide an estimate of the resources required for the successful implementation of AI solutions in administrative processes.

Cochaired by Bob Graham, assistant vice president, ERP, CRM & Integration in IS&T, and Kelly Lockard, assistant vice president, Continuous Improvement & Data Analytics, the task force comprises interdisciplinary members, including representatives from different administrative departments, and IT professionals. This diverse composition will ensure a holistic approach to addressing the challenges and opportunities associated with AI in administrative processes.

The task force is expected to submit initial findings by June 30, 2024.

Sincerely,

Kenneth W. Freeman
President Ad Interim

Boston University AI Administrative Processes Task Force

Cochairs:
Bob Graham, Assistant Vice President, ERP, CRM & Integration, IS&T
Kelly Lockard, Assistant Vice President, Continuous Improvement & Data Analytics
Members:
David Bergeron-Keefe, Assistant Director, Content & Search Engine Optimization, External Affairs
Margaret Bolter, Associate Dean, Strategic Initiatives, College of Arts & Sciences
Mumtaz Badshah Brown, Assistant Vice President, Talent Development, Human Resources
Lilly Huang, Associate General Counsel, Office of the General Counsel
Jennifer Jackson, Assistant Vice President, Digital Architecture & Technology, BU Virtual
Kris Klinger, Vice President, Auxiliary Services
Steve Koppi, Executive Director, Center for Career Development
Adam Krueckeberg, Vice Dean, School of Law
Ira Lazic, Associate Dean, Administration, School of Public Health
Linda Martin, Associate Vice President, Research Operations, Office of Research
Mark Newton, University Librarian
Ern Perez, Associate Provost, Executive Director, BUMC IT
Eddie Ramones, Director, Human Resources Information Systems, Human Resources
Kerri Saucier, Assistant Vice President, Advancement Information Strategies, Development & Alumni Relations
Staff Liaison:
Marie Vearling, Senior Consultant, Continuous Improvement & Data Analytics