

1.A. Company Name

Mindwell Technologies

1.B. Long-Term Vision Statement

1.B.1 Goals:

Mindwell Technologies aims to revolutionize workplace mental health support by leveraging AI-driven conversational agents to provide scalable, confidential, and accessible mental health assistance. The goal is to reduce workplace stress, improve employee well-being, and create a supportive corporate culture where mental health is prioritized. Over the next five years, Mindwell seeks to expand its AI capabilities, build a diverse network of licensed mental health professionals, and integrate advanced privacy measures to enhance user trust and compliance with mental health regulations.

1.B.2 Idea Origination:

The concept for Mindwell emerged from a gap identified in corporate mental health support during an AI ethics seminar in a computer science class. The increasing workplace stress and mental health crises highlighted the need for a scalable, AI-driven mental health solution that bridges the gap between self-help tools and traditional therapy. Inspired by both technological advancements and a growing corporate demand for mental health solutions, the founders saw an opportunity to integrate ethical AI practices with mental health support.

1.B.3 Purpose/Values/Mission:

Mindwell Technologies is dedicated to ethical AI implementation in mental health care, ensuring accessibility, privacy, and accuracy. The mission is to provide an AI-powered mental health support system that complements traditional therapy while maintaining the highest standards of data security and ethical AI use. Core values include:

- **Empathy & Inclusivity:** Ensuring AI-driven support is accessible to diverse user groups.
- **Privacy & Security:** Maintaining the highest standard of data protection and ethical AI practices.
- **Collaboration:** Working with licensed professionals to ensure AI interventions align with clinical best practices.
- **Continuous Improvement:** Using feedback loops to refine AI accuracy and user experience.

1.B.4 Key Questions:

1. How can Mindwell ensure that AI-driven mental health support remains ethical and unbiased across diverse populations?
2. What strategies will maintain high accuracy in AI-generated mental health recommendations while safeguarding user privacy?
3. How can Mindwell create sustainable partnerships with corporate clients while

prioritizing user well-being over profit?

1.C. Strategy with Ethical Impacts AND Ethical Safeguards

OKR 1: Improving Chatbot Screening Accuracy for Mental Health Needs

Objective: Develop a clinically sound and structured screening process in collaboration with mental health professionals to enhance the chatbot's ability to accurately assess user mental health concerns and recommend appropriate providers.

Key Result:

- Achieve an **85% provider matching accuracy** within the first six months.
- Maintain at least **90% provider matching accuracy** after one year with continuous feedback integration.

Experiment:

- Conduct controlled **user simulations** with 50-100 participants simulating various mental health concerns.
- Compare chatbot-generated problem summaries with **licensed mental health professional evaluations**.
- Implement a **feedback loop** where providers rate the chatbot's accuracy in summarizing patient concerns.

Ethical Impacts:

- **Risk of misdiagnosis or misinterpretation:** If the chatbot incorrectly assesses a mental health concern, it may delay proper care.
- **Potential AI bias:** Lack of diversity in training data could lead to inequitable provider recommendations.
- **User frustration and trust issues:** Incorrect or unclear recommendations could lead to employee dissatisfaction.

Ethical Safeguards:

- **Diverse training datasets** incorporating multiple demographics and mental health conditions.
- **Clinical expert review of chatbot outputs** before live deployment.
- **Regular audits** to assess bias in provider recommendations.

OKR 2: Enhancing Data Privacy & Security in AI Mental Health Support

Objective: Ensure user privacy by implementing advanced security measures, including **differential privacy** and **zero-retention data storage policies**.

Key Result:

- Implement **end-to-end encryption** for all chatbot-user interactions.
- Maintain **100% compliance** with HIPAA and GDPR regulations.
- Achieve **90%+ user satisfaction** in privacy and data security.

Experiment:

- Conduct **bi-annual penetration tests** to assess system vulnerabilities.
- Implement a **user opt-in model** for data st

Meeting goal: iterate group entrepreneur project, firm up OKRs, metrics, experiments

Meeting day/time: Tues 3/4 8-915am Who Attended: Marianna Belmares, Jason Russell, Luke Hagan, Benjamin Thompson

Meeting summary: Further work on OKRs, business goals and mission statement

Mission statement: Offer a chatbot as a mobile app as an additional support for common mental health challenges for corporate employees. Focused on helping user ages of 25-49, primary use is case is to find a licensed therapist in a directory. Through the interactive use of a LLM chatbot delivered in the app, company can offer employees immediate help. LLM chatbot can run screening dialogue questions to best place user with a licensed practitioner drawn from directory. Through careful safety studies, a safety approval by a clinical group would be sought to support market messaging for corporate clients. One additional language would be developed early starting with Spanish.

- OKRs : -- Chatbots for mental health:

[NO CHANGE] --- Objective: Through engaging with experts, develop a series of scripted dialogues to serve as ongoing test scenarios to assure chatbot is operating within expected norms. --- Key Result: LLM Chatbot operates within norms in scripted scenarios for common conditions. --- Objective: By engaging test users, confirm through user studies that chatbot continues to perform within safety objectives providing advice to users in studies. --- Key Result: Users report an engaging dialogue with the chatbot, and self report satisfaction with the interactions --- Objective: Through engaging clinicians in review, confirm through paid clinician review of user studies that the chatbot is operating within norms in user studies. --- Key Result: Study of clinical review of chatbot performance is within norms and satisfactory

[WILL NOT DO] --- Objective: Engage a regulator with supporting user studies that both the user and clinicians can report positive satisfaction through interactions with the chatbot: --- Key Result: Regulator solicits constrictive feedback, enabling to move forward with testing at the regulator:

[NO CHANGE] --- Objective: Get users into marketing studies to construct positive stories of use for a wider marketing campaign. --- Key Result: Marketing based study is satisfactory, giving confidence in positive testimonials to by used in user marketing --- Objective: Open user testing with a broadly appealing ad campaign with supporting studies --- Key Result: open user testing can confirm results found in smaller user studies

[ADDITION] --- Objective: Develop screening dialog with clinical experts to support placement with licensed therapists familiar with the problem and conditions and interested to help --- Key Result: LLM chatbot can be engaged with a variety of problems. A summary is automatically generated for review by clinicians as part of placement with a provider. Provider can review full text of dialogue and offer feedback to LLM chatbot team.

[ADDITION] --- Objective: App supports a provider directory, or a marketplace of providers available for each corporate client and geography --- Key Result: Starting with two to three enterprises and one geography, get ten to twenty licensed providers signed up with fully created profiles for matchmaking and user booking.

[ADDITION] --- Objective: Sell to corporate enterprises as a subscription benefit for their employees a certain number of hours of licensed provider hours placed through the provider marketplace. --- Key Result: Starting with two to three enterprises, work with one or more HR contacts to make the service available on a trial basis.

Agreed

- Jason is taking the Marketing OKRs
- Luke is taking the Safety OKRs
- Marianna to review the business plan, mission statement for another pass. Will follow up.
- Ben is will take whichever OKRs remain.

Next Steps:

Further develop metrics and experimental designs in one on one follow up meetings.

Revised the mission statement after our meeting. For each OKR, noted whether NOCHANGE, WILL NOT DO or ADDITION.

For each OKR identifying representative metrics, and how we'll measure those and any ethical considerations are the things to think about, and if you can start writing.

If you'd like to revise, simplify or change the mission statement, suggestions welcome!

Welcome to the Group-2-Computing-Responsibility-CS230-02-1-Sp25- wiki!

Group Email:

bthompson34@horizon.csueastbay.edu jrussell28@horizon.csueastbay.edu
lhagan4@horizon.csueastbay.edu mbelmares@horizon.csueastbay.edu

This is our second group meeting.

We are discussing the Case Study and the Project - Entrepreneurism & Ethics.

1. Is the proposal sufficient?
- 2.

Potential Links/Citations:

<https://www.aclu.org/cases/google-v-gonzalez-llc#:~:text=Every%20time%20one%20does%20a,light%20of%20the%20Twitter%20ruling>.

Personal Ethics Benjamin Thompson: Moderation issues

Google was taking a proactive stance by trying to moderate some of the information now. What is the content is harmful Google's Sophisticated operation of moderating information on Youtube.

Ethical Issues <https://www.supremecourt.gov/docket/docketfiles/html/public/21-1333.html>

https://www.supremecourt.gov/opinions/22pdf/21-1496_d18f.pdf

Personal Ethics

Benjamin Thompson: Moderation issues

- Google was taking a proactive stance by trying to moderate some of the information now.
- What is the content is harmful

Google's Sophisticated operation of moderating information on Youtube.

Ethical Issues

<https://www.supremecourt.gov/docket/docketfiles/html/public/21-1333.html>

https://www.supremecourt.gov/opinions/22pdf/21-1496_d18f.pdf

All were present: Benjamin, Luke, Jason, Marianna

ACM Ethics, Should we embed them in the Proposal?

Marianna is meeting with teacher at office hours today (Tuesday February 11, 2025). 11:45 AM -12:45 PM

Mari shared some links:

https://docs.google.com/presentation/d/1j8zf_mjrPPSTol-5GdH3Vc7_s_vvbppGL1IMzIDiroU/mobilepresent?pli=1&slide=id.g33370bf13ad_0_484

https://stratechery.com/2025/deepseek-faq/?utm_source=chatgpt.com

<https://www.meetup.com/silicon-valley-generative-ai/events/300239972/?notificationId=1470559610228609024&eventOrigin=notifications>

https://www.meetup.com/building-ai-together-san-francisco/events/305885733/?eventOrigin=group_similar_events

We talked about potential meetings.

The use of Github in our project.

Writing our paper, posting in a shared Google Doc and going over it Sunday.

Next Steps:

Twitter vs. Taamneh - read those dockets relevance. Read dockets from Gonzalez vs. Google.
Meet Thursday 8 PM briefly - accountability. Sunday - meet to discuss the rough draft.

We went over the group project and Marianna sent what

<https://cdn.ca9.uscourts.gov/datastore/opinions/2021/06/22/18-16700.pdf>

“Plaintiffs also provide a set of allegations specific to Google. According to plaintiffs, Google has established a system that shares revenue gained from certain advertisements on YouTube with users who posted the videos watched with the advertisement. As part of that system, Google allegedly reviews and approves certain videos before Google permits ads to accompany that video. Plaintiffs allege that Google has reviewed and approved at least some ISIS videos under that system, thereby sharing some amount of revenue with ISIS. ”

“————— 4Plaintiffs also raised other claims, including that defendants were directly liable for having provided material support to ISIS. See, e.g., 18 U. S. C. §§2333(a), 2339A, 2339B, 2339C. The District Court dismissed those claims as well, and plaintiffs did not appeal them. 5The ATA defines “international terrorism” to mean “activities that— “(A) involve violent acts or acts dangerous to human life that are a violation of the criminal laws of the United States or of any State, or that would be a criminal violation if committed within the jurisdiction of the United States or of any State; “(B) appear to be intended— “(i) to intimidate or coerce a civilian population; “(ii) to influence the policy of a government by intimidation or coercion; or “(iii) to affect the conduct of a government by mass destruction, assassination, or kidnapping; and “

Meeting goal: review Gonzalez/Tamneeh case details, continue group entrepreneur project
Meeting day/time: Tues 2/18 8-915am Who Attended: Marianna Belmares, Jason Russell, Luke Hagan, Benjamin Thompson

Meeting summary: Reviewed Gonzalez/Tamneeh particulars, discussed how to wrap up move to docs, presentation. Marianna has text to review. Ben has text to write. Asked for sentences from group for Case Analysis, & Ethics sections.

- Agreed to pull together notes Sunday Discussed several group entrepreneur angles and ethics, mentioning potential OKRs
- Rejected several concepts as minimal to no interest: legal discovery, state tools, inauthentic social interactions
- Narrowing down to: LLM/RL Brand Management, Ad Recommenders Brand Management
- OKRs Scenario A: -- LLM/RL Brand Reputation Monitoring --- Objective: Through browser sampling, monitor how LLM chatbots are discussing trademarked brands for rights holders --- Key Result: Automated production of examples of poor brand placements for rights holders --- Sales process for a monitoring product for brand reputation in the LLM age.
- OKRs Scenario B: -- Recommenders Brand Management -- Content placement Brand Reputation Monitoring --- Objective: Through browser sampling, monitor the recommended content next to branded content recommendations. Map adjacent content for reputational risks --- Key Result: Automated production of examples of bad content neighbors for rights holders --- Sales process for a monitoring prodjct for brand reputation of content within recommenders.

Next Steps:

Further develop consensus on group entrepreneurial approaches, techniques, objectives, key results. Weekend deliver more text for group ethics project, move document forward

with Mari

March 11 - Individual contribution due.

Sentences about Gonzalez Case - Benjamin Thompson, Luke

- Stop it from giving physical medical advice
- There is a lot of potential for ethics
- OKRs

Meeting goal: continue group entrepreneur project Meeting day/time: Tues 2/25 8-915am Who Attended: Marianna Belmares, Jason Russell, Luke Hagan, Benjamin Thompson

- OKRs : -- Chatbots for mental health: --- Objective: Through engaging with experts, develop a series of scripted dialogues to serve as ongoing test scenarios to assure chatbot is operating within expected norms. --- Key Result: LLM Chatbot operates within norms in scripted scenarios for common conditions. --- Objective: By engaging test users, confirm through user studies that chatbot continues to perform within safety objectives providing advice to users in studies. --- Key Result: Users report an engaging dialogue with the chatbot, and self report satisfaction with the interactions --- Objective: Through engaging clinicians in review, confirm through paid clinician review of user studies that the chatbot is operating within norms in user studies. --- Key Result: Study of clinical review of chatbot performance is within norms and satisfactory --- Objective: Engage a regulator with supporting user studies that both the user and clinicians can report positive satisfaction through interactions with the chatbot: --- Key Result: Regulator solicits constrictive feedback, enabling to move forward with testing at the regulator: --- Objective: Get users into marketing studies to construct positive stories of use for a wider marketing campaign. --- Key Result: Marketing based study is satisfactory, giving confidence in positive testimonials to be used in user marketing --- Objective: Open user testing with a broadly appealing ad campaign with supporting studies --- Key Result: open user testing can confirm results found in smaller user studies

Further develop consensus on OKRs and experimental design of each Weekend discussion of substance of experimental designs.

Key target groups:

Adult 27 - 50
Working

Discussing OKRs with group What benefits the business?

Retention Rate: User Satisfaction:

Compliance with HIPPA AND GDPR:

Expand User Base - Accessibility:

Ensure Ethical AI Decision Making

Measure use of app features for specific conditions

- Work with existing clinicians
- But no active therapy with human to avoid FDA regulation
- Chatbot - facilitate scheduling, talking about work
- Offering wellness plan
- Offered by employees
- Interact chatbot with therapy
- Immediate help
- Comparisons with current mental wellness apps

Comparisons with competitive apps Talkspace Therapy Headspace

Harmonization of Entrepreneurial Strategy with Ethical Impacts & Safeguards.

Objective 1: Enhance User Engagement with Therapy Services KR1: Achieve a % retention rate of users after three months. KR2: Increase the average session completion rate to % within six months. KR3: Ensure that % of users interact with AI therapy at least once per week.

Objective 2: Improve the Effectiveness of AI-Assisted Therapy KR1: Achieve % user satisfaction rating based on post-session surveys. KR2: Reduce the average session dropout rate to below %. KR3: Ensure that at least % of users report symptom improvement after six weeks.

Objective 3: Strengthen Data Security and Privacy Compliance KR1: Maintain % compliance with HIPAA and GDPR regulations. KR2: Conduct biannual third-party security audits with a passing score of % or higher. KR3: Implement end-to-end encryption for % of user communications and stored data.

Objective 4: Expand User Base and Accessibility KR1: Reach active monthly users within the first year. KR2: Provide multi-language support and onboard at least non-English-speaking users. KR3: Partner with three major insurance providers to expand affordability.

Objective 5: Ensure Ethical AI Decision-Making KR1: Conduct quarterly bias audits on AI therapy recommendations with a % discrimination rate. KR2: Ensure that % of AI-generated

suggestions align with licensed therapist recommendations. KR3: Implement an AI transparency dashboard accessible to users and regu

Luke, Marianna, Benjamin, Jason

We discussed the app, went over the guidelines of the project. Repeated the OKR

We discussed final additions to our draft so far.

Analysis of Gonzalez v. Google LLC Group 2 Members: Marianna Belmares, Benjamin Thompson, Luke Hagan, Jason Russell

Part 1: Case Synopsis Stakeholders: Plaintiffs: The Gonzalez family (parents of Nohemi Gonzalez). Defendant: Google LLC (owner of YouTube). Legal Entities: The Supreme Court of the United States, lower courts that reviewed the case. Public & Governments: Legislators, policy advocates, and the global tech industry. Case Description: Gonzalez v. Google LLC, 598 U.S. 617 (2023) was a landmark case brought before the U.S. Supreme Court of the United States when the family of Nohemi Gonzalez petitioned for writ of certiorari after their initial lawsuit, filed against Google in 2016, ruled in favor of Google by the U.S. Court of Appeals for the Ninth Circuit in October of 2020 [1]. The case emerged after Nohemi Gonzalez, a U.S. citizen, was killed in a coordinated terrorist attack by the group ISIS on November 13, 2015 while she was dining with friends at a cafe in Paris. In response to her death, Gonzalez' family filed a lawsuit against Google, the parent company of Google, alleging that the platform's algorithm facilitated the spread of ISIS propaganda, contributing to the operational planning of coordinated attacks like the one in Paris [2]. The plaintiffs originally sought damages under the Anti-Terrorism Act (ATA) against Google, Twitter and Facebook under the basis that the platforms allowed the terrorist group, ISIS, to post harmful content, communicating terrorist messages, radicalizing new recruits and furthering terrorist missions abroad [3]. In October of 2020, the U.S. Court of Appeals for the Ninth Circuit stated that Section 230 of Communications Decency Act shielded social media platforms from liability for third-party content, which included algorithmic recommendations [2]. When the Gonzalez family petitioned the U.S. Supreme Court to review the case, they challenged whether Section 230 should shield platforms when their algorithms promoted harmful content. The U.S. Supreme Court's decision to hear the case was a major challenge to Section 230, questioning whether tech companies should be held liable for algorithm-driven content recommendations. A ruling against Google could have reshaped platform accountability, impacting content moderation, free speech, and the legal framework governing Big Tech [4]. Outcome: In May 2023, the U.S. Supreme Court dismissed the case, ruling that the plaintiffs failed to establish direct liability under the Anti-Terrorism Act. The Court did not directly address Section 230, but instead remanded the case to the Ninth Circuit for reconsideration in light of its decision in Twitter, Inc v. Taamneh, where the Supreme Court ruled that social media platforms could not be held liable for terrorist attacks unless plaintiffs can prove that platforms knowingly and substantially aided in a specific act of terrorism [5]. The significance of this outcome reinforced the broad legal protections for Big Tech companies under existing laws and the avoidance of direct ruling on Section 230 preserved the status quo, preventing major shifts in platform liability that could have drastically altered how online services handle content recommendations and shielded massive corporations like Google, Twitter and Meta from increased liability risks.

Part 2: Personal Ethics Should Google bear responsibility for algorithmic amplification of harmful content on its Youtube platform? Marianna Belmares: I personally believe that Google should bear some responsibility for algorithmic amplification of harmful content on its Youtube platform. It's company mission is "to organize the world's information and make it universally accessible and useful," there must be a clear ethical boundary when free access to information compromises human safety. Algorithms are not neutral, they are designed to maximize engagement, and when they actively push harmful content, the company has a moral obligation to intervene. Accountability and corporate responsibility

needs to be prioritized and Google's obligation to the public and its involvement in discourse is to mitigate potential harm through improved content moderation, developing more responsible recommendation models and greater transparency with user awareness and how Google's platforms influence content consumption and recommendation. Part 3: Professional Ethics The ACM Code of Ethics was designed to guide ethical conduct of computing professionals, future practitioners or anyone who uses computing in a way that impacts others. The Code serves as a foundation for addressing violations when they occur and outlines principles as statements of responsibility, emphasizing that the public good should always be the foremost concern [6].

Violations of ACM'S General Ethical Principles.

1.1 Contribute to society and to human well-being, acknowledging that all people are stakeholders in computing. ACM states that "an essential aim of computing professionals is to minimize the negative consequences of computing" and "should consider whether the results of their efforts will be used in socially responsible ways," [6]. Google has violated this first code of ethics by providing an open platform that enabled the widespread dissemination of harmful content. The lack of sufficient safeguards against the amplification of extremist, misleading and dangerous material through its recommendation algorithms raises serious ethical concerns.

1.2 Avoid Harm. Google violated the second code of ethics by disseminating content through Youtube which caused physical and mental injury. These videos were used to spread propaganda, recruit new members and radicalize individuals and included widely circulated videos like "There's No Life Without Jihad," which featured three British ISIS fighters glamorizing terrorism and encouraging viewers to join ISIS. Although YouTube made efforts to remove the video, it reappeared multiple times. Other videos included the leader of ISIS, Abu Bakr al-Baghdadi, who used YouTube to call on Sunni youths worldwide to join the fight, framing it as a religious duty and violence and execution videos which were designed to intimidate opponents and attract recruits who were drawn to violence [7].

Part 4: Comparison

Part 5: References SCOTUSblog. Gonzalez v. Google LLC, Supreme Court of the United States, 2022. Available: <https://www.scotusblog.com/case-files/cases/gonzalez-v-google-llc/>. [Accessed: 15-Feb-2025]. U.S. Court of Appeals for the Ninth Circuit, Gonzalez v. Google LLC, No. 18-16700, June 22, 2021. Available: <https://cdn.ca9.uscourts.gov/datastore/opinions/2021/06/22/18-16700.pdf>. [Accessed: 15-Feb-2025]. Justia Law. Gonzalez v. Google LLC, No. 18-16700, U.S. Court of Appeals for the Ninth Circuit, June 22, 2021. Available: <https://law.justia.com/cases/federal/appellate-courts/ca9/18-16700/18-16700-2021-06-22.html>. [Accessed: 15-Feb-2025]. Covington & Burling LLP, "The U.S. Supreme Court punts on Section 230 in Gonzalez v. Google LLC," May 2023. Available: <https://www.cov.com/en/news-and-insights/insights/2023/05/the-us-supreme-court-punts-on-section-230-in-gonzalez-v-google-llc>. [Accessed: 15-Feb-2025]. Supreme Court Ruling: Gonzalez v. Google LLC, [Online]. Available: https://www.supremecourt.gov/opinions/22pdf/21-1333_6j7a.pdf. ACM Code of Ethics, Association for Computing Machinery, [Online]. Available: <https://www.acm.org/code-of-ethics>. I. Awan, "Cyber-Extremism: Isis and the Power of Social Media," Society, vol. 54, no. 2, pp. 138–149, Mar. 2017. Available: <https://link.springer.com/article/10.1007/s12115-017-0114-0>. [Accessed: Feb. 15, 2025].

N. Persily and J. Tucker, "Social Media and Democracy: The State of the Field," Cambridge

University Press, 2020. J. Zittrain, "The Future of Internet Regulation," Harvard Law Review, vol. 132, no. 5, pp. 1234-1260, 2019. B. Ghosh, "Algorithms and Bias: The Ethics of AI Curation," AI & Society, vol. 35, no. 3, pp. 501-516, 2021. Oyez. Gonzalez v. Google LLC, No. 21-1333, Supreme Court of the United States, 2022. Available: <https://www.oyez.org/cases/2022/21-1333>. [Accessed: 15-Feb-2025]. Supreme Court of the United States, Petition for a Writ of Certiorari: Gonzalez v. Google LLC, No. 21-1333, Apr. 4, 2022. Available: https://www.supremecourt.gov/DocketPDF/21/21-1333/220254/20220404211548101_GonzalezPetPDF.pdf. [Accessed: 15-Feb-2025]. J. D. McKinnon, "Google case heads to Supreme Court with powerful internet shield law at stake," The Wall Street Journal, Feb. 20, 2023. Available: <https://www.wsj.com/articles/google-case-heads-to-supreme-court-with-powerful-internet-shield-law-at-stake-e548e241>. [Accessed: Feb. 15, 2025].