Mindwell Technologies

Jump to [PDF version of report](#)

Jump to [MP4 presentation of report](#)

Jump to [YouTube Presentation](#)

Project Report

Jump to any wiki Section, or sub-section below

1 Ethical Business Plan

1.A. Company Name

1.B. Long-Term Vision Statement

1.B.1 Goals

1. Ethical Business Plan

1.A. Company Name

Mindwell Technologies

1.B. Long-Term Vision Statement

1.B.1 Goals:

Mindwell strives to revolutionize workplace mental health support through the use of AI-based conversational agents to provide scalable, confidential and accessible mental health assistance. Mindwell seeks to deliver accessible health care to employees through an AI-powered mobile app and web based platform which is designed to support common mental health challenges such as stress, communication issues, trauma, including PTSD and burnout.

Mindwell will offer its' users a conversational Large Language Model (LLM) chatbot that screens users' concerns and offers self-help guidance or connects them to licensed therapists, operating 24/7 to ensure employees can always access mental health support. Mindwell helps HR leaders evaluate service impact and effectively reduce employee burnout and absenteeism, demonstrating ROI, through detailed dashboard and utilization reports, on mental health investments.

Feedback from users and licensed providers is continuously integrated into Mindwell's LLM development to enhance problem summaries, refine provider recommendations, and improve conversation quality delivering the highest-quality user experience. At the core of Mindwell's mission is a steadfast commitment to ethical AI standards and responsible data practices, including HIPAA compliance, robust data encryption, differential privacy, and the careful handling of sensitive user information. Comprehensive safeguards are consistently implemented to prevent bias, ensure provider quality, and uphold user privacy.

1.B.2 Idea Origination:

The concept for Mindwell emerged from a gap identified in corporate mental health support and the emergence of AI in all facets of industry. 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 Quality to Facilitate High-Accuracy Therapist Placement

1.C.1.1 Objective:

Develop a clinically sound and structured screening process in collaboration with licensed mental health professionals to improve Mindwell's AI chatbot's ability to accurately identify user mental health concerns and recommend appropriate, pre-vetted providers. The AI should generate concise and contextually relevant problem summaries that enable licensed providers to quickly understand user needs and determine treatment fit. This objective aims to streamline the matching process, enhance user satisfaction, reduce provider workload, and build trust among users, HR clients, and mental health professionals.

1.C.1.1 Key Result:

Mindwell assists users in dealing with a wide spectrum of mental health issues, such as anxiety, workplace stress, trauma, and burnout. Through thoughtful, contextualized conversations, Mindwell extracts important elements of a user's experience and translates them into readable, concise summaries for clinical review. With this knowledge, Mindwell is able to suggest a curated list of licensed mental health providers whose practice addresses the user's needs in a more timely manner.

Mindwell's clinical network is critical in shaping this process. Licensed mental health providers review the chat summaries to ensure accuracy and provide individualized treatment plans based on the individual's needs. This feedback allows Mindwell to increase its ability to recognize the nuances of user concerns, improve the quality of the conversation, and improve provider recommendations. This collaborative process establishes an ongoing feedback loop that is always improving - ideally leading to better provider matches, richer user experiences, and increasingly effective mental health support

.

1.C.1.2 Metric(s) with Experiment(s):

Metric 1: Chatbot-to-Provider Matching Accuracy (Primary Outcome Metric)

This metric measures the accuracy of the chatbot's provider recommendations by assessing how well the chatbot can match users with the most suitable licensed mental health provider based on their expressed problem. Accuracy will be measured as the percentage of chatbot-recommended providers that both align with the user's described concern and are approved by the licensed provider after reviewing the summary. The target of this metric is to achieve 85%+ provider matching accuracy within the first 6 months of product launch and maintain at least 90% matching accuracy after one year with continuous feedback from providers.

Experiment to measure Chatbot-to-Provider Matching Accuracy:

Purpose:
To test whether the LLM chatbot can accurately identify user problems and match them to a provider whose specialty aligns with the user's expressed need.

Setup: To test the chatbot's ability to address real-world mental health concerns, Mindwell will recruit 50–100 participants to simulate employee challenges such as stress, trauma, and conflict using pre-scripted cases. Participants will describe their concerns in their own words, allowing the chatbot to generate a concise problem summary and a tailored list of 5–8 licensed providers aligned with their needs.

Step 1: To validate the chatbot's provider recommendations, each chat log will be reviewed by an independent licensed mental health professional unaffiliated with Mindwell. The expert will select a provider based on the transcript, and the chatbot's top 5–8 suggestions will be compared to this choice. A match is considered accurate if the expert's selection appears in the chatbot's list, enabling a quantitative measure of alignment with clinical judgment.

Step 2: To calculate matching accuracy, a formula, 'The number of chatbot matches aligned with expert's choice' against the 'Total number of user cases' is applied with an acceptable threshold of 85% accuracy of top 5 provider matches that align with an expert's choice.

Step 3: After a user's first session, the matched provider will rate the chatbot-generated summary on a 1–5 scale and may offer qualitative feedback if key concerns were missed or misinterpreted. This input will be reviewed and used to refine the chatbot through ongoing training. The goal is to achieve over 85% recommendation accuracy initially and surpass 90% within the first year. Positive provider feedback will validate summary quality and ease clinician workload. Both quantitative ratings and qualitative evaluations will be collected, and if accuracy drops below 85%, the product team will prioritize retraining. Continuous provider feedback will guide quality improvements and support effective caseload management [1].

Metric 2: Problem Summary Quality (User-to-Provider Information Transfer)

Definition:
This metric assesses how clearly and accurately the chatbot's summary conveys the user's concern, emotional tone, and specific issue category, ensuring it is useful for provider review.

Experiment to Measure Problem Summary Quality:

Purpose: To measure how accurately the chatbot captures the user's concern and translates it into a helpful summary for the provider.

Setup: After each user-provider session, providers will receive the chatbot's problem summary and rate its clarity, relevance, and brevity on a 1–5 scale, evaluating how clearly and concisely it captures the user's core concern.

Data Collection: Takes the average scores for each session and tracks the percentage of sessions that receive a 4.5+ score.

Metric 3: User Satisfaction with Provider Match

Purpose: This metric gathers feedback from users, using surveys, following their first session with a matched provider. The survey asks the following questions: Did the provider feel like an adequate fit for their issue? Did the chatbot understand your concern? Would you continue to use Mindwell?

The target for this metric is: at least 80% of users report that the recommended provider was a "good fit," and at least 85% of users report they felt understood by the chatbot [2].

Experiment to Measure User Satisfaction:

Purpose: To directly measure user satisfaction after their first session with a recommended provider.

Users will complete a short post-session survey rating provider fit, chatbot accuracy, and overall helpfulness. Responses include yes/no questions and 1–5 scale ratings. Data will track high satisfaction (4–5 ratings) and willingness to reuse the platform. Expected outcomes include 80%+ reporting a good provider match, 85%+ feeling understood by the chatbot, and 75%+ willing to use Mindwell again. If satisfaction or accuracy falls below target, the LLM will be retrained to improve understanding of user concerns and emotional cues.

1.C.1.3 Ethical Impacts/Issues:

Mindwell's chatbot raises ethical concerns around data privacy, AI bias, and misrepresentation. Collecting sensitive mental health data poses risks if improperly handled, while biased training data can lead to poor provider matches, especially for underrepresented users. A notable example is BetterHelp, which faced backlash in 2023 for sharing user data without consent, highlighting the consequences of privacy violations.

Key risks include data breaches, biased assessments, and inaccurate summaries that may misdirect care. These issues impact users, providers, developers, and the company's reputation. To prevent harm, Mindwell must enforce strong privacy safeguards, ensure diverse training data, and conduct ongoing audits to maintain user trust and ethical AI

standards.

1.C.1.4 Ethical Safeguards:

Mindwell will implement differential privacy by adding random noise to data, ensuring no individual can be identified. All data will be stored on HIPAA-compliant cloud infrastructure (e.g., AWS, Google Cloud) with encryption, access controls, and automatic deletion. A Data Privacy Officer will oversee audits, and user privacy satisfaction will be measured through regular surveys and bi-annual penetration tests.

To prevent algorithmic bias, Mindwell will use diverse training data representing various ethnic, linguistic, and socioeconomic groups. A Bias Mitigation Task Force (BMTF) will evaluate outputs, and equity will be tracked by comparing provider match rates across demographics. Disparity should remain under 5%.

Mindwell will establish a provider feedback system to report inaccurate summaries or mismatches. Using human-in-the-loop (HITL) processes, therapists will help retrain the model. A feedback portal and quarterly reviews will support accuracy improvements, with goals to boost provider satisfaction by 20% and reduce misclassifications by 15% annually.

Mindwell's success depends on its ability to deliver accurate, efficient, and ethical mental health screening through AI. By combining structured clinical collaboration, diverse data curation, real-time feedback from providers, and robust privacy protocols, this OKR ensures Mindwell can scale responsibly while meeting the mental health needs of corporate workforces [3].

1.C.2.1 OKR 2: Chatbot Safety for Mental Health Provider Matching

Objective: By engaging with experts in mental health, we should develop a set of boundaries where the chatbot should be able to ask screening questions without acting as an impromptu therapist. In order to be able to match the user with a suitable therapist the chatbot would need to ask questions to determine the nature of the user's issues, but it shouldn't delve too deeply into the specifics where a licensed therapist would be more suitable.

1.C.2.1 Key Result: A chatbot, when put through a set of scenarios, is able to determine a suitable stopping point where it has attained the minimum amount of data to give a suitable recommendation for a therapist. We should be able to verify that 95% of total recorded interactions with the chatbot are considered safe.

1.C.2.2 Experiment:

Preliminary Testing:

Collect a sample of chatbot transcripts from a group of test users who would carry out a simulation of both scripted scenarios derived from the advice of licensed practitioners and unscripted scenarios where test users would role play common scenarios that showcase various sources of stress and anxiety among working adults.

1.C.2.2. Metrics: Evaluate the chatbot's interaction with the user using the following:

- Was the chatbot able to determine an appropriate stopping point where it has attained

the minimum amount of information for a suitable recommendation?
- Has the chatbot avoided asking overly specific questions that prompt users to divulge an inappropriate amount of information?

Target: Over 95% of interactions can be applied to the above evaluation.

Chat Log Assessment:

Collect a sample of transcripts from users to be graded by licensed practitioners who will determine whether the conversations are considered to be appropriate for an interaction with a language model.

Metrics: Determine overall grade of the number of inappropriate interactions in each sample, as assessed by the experts.

Target: The overall number of safe interactions as determined by experts is above 95%.

1.C.2.3. Ethical Impacts:

- **Data Handling:** The users must be informed that their chat logs are being collected and how it is being handled by our business.
- **Possible Data Breaches:** Any chat logs that are being stored can be subject to a possible data breach.
- **Overstepping its Boundaries:** Even if all the experiments are considered to be successful, there is still a non-zero chance of the chat bot overstepping its boundaries and acting as a replacement for a therapist.

1.C.2.4. Ethical Safeguards:

- **Disclaimers:** In order to be transparent about the usage of their data, we will post disclaimers around the app that state that any information posted in the chat logs would be stored and they will only be seen by licensed practitioners. We could work with the practitioners themselves to add verification to that claim. We could measure its effectiveness by reading reviews on the app store.
- **Opt-out of Data Collection:** Since we are a business stationed in California, we will comply with the California Consumer Privacy Act, which ensures the right to opt-out of sharing personal information. [4] We would implement it in a way where we wouldn't need to verify whether or not you are in California, so that any user can choose to opt-out as they please.
- **Purging Chat Logs:** In order to minimize the impact of any data breaches, we will purge all chat logs after a limited period of time. We could measure its effectiveness by keeping track of the number of chat logs we have purged.
- **HIPPA Compliant Cloud Provider:** In order to lower the possibility of data breaches, we will be retaining our computing and storage in a cloud provider that is HIPPA compliant.
- **Improving the Chatbot:** In order to minimize the number of cases where the chatbot oversteps its boundaries, we will work with licensed experts to include more scenarios. We may not be able to completely get rid of the problem, we can try.

1.C.3.1. OKR 3: Patient Market Fit and User Growth

Objective: -Transform mental health support by making professional therapy more

accessible through our AI-powered chatbot.

Key Result: Make 50,000 successful user-therapist connections within the first year by using AI-driven matching, healthcare partnerships, and targeted outreach.

1.C.3.2. Experiment: To achieve 50,000 successful client connections, we will largely use ads. We will utilize paid promotions and direct user feedback to scale the company. These paid promotions will come in the form of:

- Paid contracts to post our company as ads on websites
- Pay to create and air TV commercials displaying our product and reviews

Additionally, surveys will provide insight into the user's experience and satisfaction. After a patient has completed a meeting with a therapist through the use of our app, we will ask the following questions:

- "Were you satisfied with the process?"
- "What can we improve on to make your experience better next time?"
- "Did your doctor/therapist help solve your problem?"

The responses will allow us to make continuous improvements, ensuring our app gives the best possible results to all of its users. To directly keep track of every user, our database will hold:

- Records of everyone's name, address, DOB, and any other legal information.
- Keep track of every completed session
- Keep track of the number of surveys completed
- Keep track of new user signups on our app

This way, we will be able to keep track of and sort every user of our app.

1.C.3.3. Ethical Impacts:

Many ethical concerns will arise with the creation of our app. Key issues deal with privacy risks, AI biases in therapist matching, misdiagnosis, and dependency on automated systems. Users will be sharing personal mental health struggles with the chatbot, which usually isn't something you want to share. For example, a data breach or improper data handling could expose very sensitive patient information, leading to psychological harm and a lack of trust in our service. A real-world case highlighting this risk is the Facebook-Cambridge Analytica scandal, where users' data was harvested without consent and misused for political influence [5].

If our chatbot's data is improperly stored or shared with third parties for marketing purposes, it will violate user trust and ethical AI principles. Furthermore, about targeted outreach, ethical concerns may arise regarding gathering user information to target new users. All information sent and stored on our cloud should be private, and could raise ethical concerns if we use that information, and try to market those struggles to get people to relate.

Another ethical issue is bias in AI-driven therapist matching. If the chatbot's machine learning model is trained on biased datasets, it could recommend certain therapists over others, excluding a more qualified therapist for a specific situation. If our chatbot assigns therapists based on flawed algorithms, some users may not receive appropriate care. A

third ethical risk is over-reliance on chatbot-generated advice, possibly leading to receiving false medical help. If a user experiencing a severe mental health crisis receives the wrong or not enough guidance from the chatbot or is placed on a long waiting list for therapy, it could result in harmful delays in treatment.

1.C.3.4. Ethical Safeguards:

To ensure our customers don't fall victim to the many ethical risks, we will put safeguards in place. One key safeguard is strong encryption and strict data access controls to protect sensitive user information. We will work with cybersecurity experts, data privacy lawyers, healthcare IT professionals, ensure HIPAA compliance and conduct regular security audits to limit data exposure [6].

We will constantly monitor for security attacks and gather user feedback on data security. To address AI bias in therapist matching, will will train the AI models on diverse datasets to eliminate as much bias as we can. AI ethics researchers, data scientists, and mental health professionals will work together to track user demographics in relation to therapist assignments, collect user feedback, and monitor therapist mismatch complaints to strengthen our AI recommendations.

Additionally, we will prevent long wait times by adding a level of emergency feature that will prioritize patients in need of the most immediate help. Measuring the effectiveness will involve tracking crisis escalation rates, response times, and analyzing incident reports. By creating and implementing these safeguards, we can ensure our chatbot remains safe, unbiased, and an ethical tool for mental health support.

1.C.4.1 OKR 4: Provider Market Fit and Corporate Customer Growth

Objective: Mindwell offers a provider directory with booking availability for each corporate client in multiple geographies.

Key Result: For each metro area, *achieve a consistent weekly rate of appointment booking for providers* by Mindwell users. Gradually increase the rate of appointment booking per provider to meet their desired business goals, and the *maintain metro provider pool size with availability even with provider churn*. Monitor qualitative aspects through random sampling and review.

1.C.4.2 Metrics with Experimentation

To support provider pool growth, Mindwell will track two key metrics: the *weekly rate of provider appointment bookings* and the size of the *available provider pool by metro and timezone*. The first metric ensures providers receive at least one booking attempt per week to prevent disengagement or churn. The second monitors whether each metro—or its fallback timezone—offers 3–5 providers with open slots during peak scheduling times.

We will run experiments to adjust booking distribution fairly across providers by analyzing booking patterns over recent weeks and similar holiday periods. This optimization aims to *balance demand and supply, ensuring equitable appointment allocation*. A second experiment will focus on *maintaining availability and reducing provider churn* by intervening when booking activity declines. By keeping buffer availability, following up with providers who disable bookings, and analyzing churn, Mindwell aims to sustain a healthy provider

ecosystem that reliably meets user demand.

1.C.4.3 Ethical Impacts/Issues

Operating a provider network through encrypted video conferencing with Mindwell users introduces several ethical risks. Although Mindwell does not explicitly represent providers, it is responsible for the quality of user-provider interactions and the provider experience when receiving bookings. Offering quality ratings for providers or users poses additional ethical risks, as this emerging legal area (e.g., Liu vs. Uber [7]) raises concerns about potential discrimination lawsuits. Mindwell is monitoring these developments carefully.

The core stakeholders are the employee users, providers, and HR Benefits managers who purchase service packages. Employee users face high privacy risk due to the sharing of protected health information, though financial and conflict-of-interest risks are low. Mindwell mitigates privacy risks through strict data controls and HIPAA compliance. HR Benefits managers face high financial risk, as they invest in credits for chatbot and provider services, but privacy and conflict risks are minimal. Mindwell supports customers through free trials, discounts, and adjustable usage plans.

1.C.4.4 Ethical Safeguards

In order to design our ethical safeguards in detail we have an internal team of clinical experts that we pay that offers the team insight and guidance on ethical issues. In order to get some distance from the problem we convene an external advisory panel of clinical experts that we use to see the larger picture. Through detailed reviews of data controls, experimental designs, and sampled data, the internal team of clinical experts and data scientists work together to establish and maintain our ethical safegaurds around Mindwells continuing operations.

2. Cultural Policy

2.A. Core Values

At Mindwell Technologies, our core values align at the intersection of empathy, innovation and ethical responsibility. We would like to be seen as a mental health company that leads with heart and builds with intention. Our company values trust, transparency and transformative care. Our values articulate our purpose: to radically transform workplace mental health with the ethical use of AI and support real human needs.

We believe in *empathy* and *inclusivity*. Our AI systems are able to adapt to a wide spectrum of users including diverse languages, cultures, socio-economic backgrounds and emotional experiences. We firmly believe that every user should feel heard, and their needs are well served by our providers. Privacy and security are unquestionably non-negotiable. From day one, we will uphold the highest standards of *data protection* with a HIPAA-compliant technology infrastructure and a zero-retention policy. Because clinical collaboration is integrated into our systems, our AI technologies will not be siloed and isolated tools but rather be built and validated in unison with licensed professionals and *evidence-based practices*. Transparency is critical to our culture: we consistently explain how our AI models work, how we use data to provide recommendations. Finally, we value *continuous improvement* at every level. Our development cycle will utilize exhaustive feedback loops to help align our practices and systems to the needs of users and the ethical expectations of institutional stakeholders.

2.B. Motivation

What drives Mindwell Technologies is a passion for tackling real problems with ethical, human-centered technology. We want to build tools that help people feel heard, cared for, and connected to resources that are meaningful to them. Our first focus is on high-stress work settings that may unintentionally discourage access to mental health support. Our users work where mental health is at times stigmatized by individuals, and at other times cultural trends. We envision a service where AI is not a replacement for human connection, but is a bridge to it. A convenient application that increases access, improves outcomes, and enhances the therapeutic process.

We fear is what happens if we don't get it right. We fear AI systems that misinterpret vulnerable users, and reinforce bias. We fear a reduction in trust if privacy is broken. We fear technology that oversimplifies complex mental health issues. Fear puts us in a place of caution and rigor. It drives us to provide clinical oversight, to train on diverse datasets, and to center our users over profit. We fear complacency over failure. In mental health errors do not mean bugs, they mean missed opportunities for healing. Our culture values responsible risk-taking, ethical design, and a mindset of reflection and improvement.

2.C. Summary

Compassionate AI Built for Human Wellness

3. Ethics Policy

3.A. Core Items

Mindwell Technologies is committed to applying ethical principles at every stage of our AI-enabled mental health platform. Our policies reflect how we collect, process, and act on sensitive information. Our focus on user safety and clinical appropriateness connects user clients with providers, and satisfies payers.

1. Privacy and Data Ownership Within Mindwell's platform, all data shared by the user remains confidential and owned by the user. Mindwell does not store personally identifiable information (PII) without user consent. In addition, we employ differential privacy and encrypt all interactions with our chatbot. We regularly assure users how their sensitive data will be handled. Limited data retention can be modified by a user facing opt-in or opt-out selection. When a user is opt-in, recent data is used for training our AI model with privacy preserving anonymized data [8].

2. Algorithm Fairness and Bias Mitigation

Mindwell's AI systems undergo stringent audits to mitigate bias and promote equitable outcomes. We curate our training data to include the full spectrum of ethnicity, language, and socioeconomic variability. We employ adversarial testing to ensure fairness in provider recommendations. Our Bias Mitigation Task Force assures cultural sensitivities are well served by including ethnically and socioeconomically diverse team members.

3. Transparency and Explainability

Our large language model generative AI is focused on creating summaries that are accurate and explainable to the user and clinician. Providers receive access to conversations in full transcript form, and a summary is generated by AI to confirm that the provider aligns with the user client need. Users are advised when generative AI is operating, and how AI recommendations are derived.

4. Clinical Validation and Human Oversight

Mindwell uses human-in-the-loop (HITL) review after automatic classification, escalating high-risk use cases directly to a human clinician. Our chatbot screening process is developed with licensed therapists, and clinician feedback is used continuously for improvements to our AI solutions [9], [10].

5. Corporate Integrity and Accountable monetization

Mindwell Technologies has a clear separation between user data and corporate profit. Mindwell Technologies never sells user data. Our intention is to align monetization strategy with ethical service delivery by offering access to prepaid provider sessions for corporate customers.

6. Regulatory Compliance and Management of Safety

Mindwell abides by HIPAA, and GDPR regulations, and is currently engaged in a third-party safety validation process. Our security practices include conducting penetration testing of

our platform every six months. We use zero-retention data usage practices. We continuously monitor our platform to follow best practice guidance of relevant regulatory agencies such as HHS and NIST [11], [12], [13].

3.B. Ethics Board

1. Dr. Thomas Insel – Formerly, Director of the National Institute of Mental Health (NIMH)

Dr. Thomas Insel is a psychiatrist and neuroscientist who was the Director of the National Institute of Mental Health for over a decade. Previously Dr. Insel co-founded Mindstrong Health, a digital mental health start-up to build and create clinical neuroscience with mobile technology. He is a leading proponent for the ethical use of technology for psychiatry, and is particularly interested in solutions that provide access to mental health care while maintaining the integrity of clinical research. Because of his unique background in clinical care and technology in mental health care, Dr Insel is well positioned to be a member of our ethics board, and advise Mindwell on AI-powered mental health care [14].

Dr. Thomas Insel brings a unique blend of clinical expertise, technological insight, and ethical leadership to support Mindwell's mission of transforming workplace mental health through an AI-driven, compassionate care model. Dr. Insel has been at the forefront of integrating neuroscience with mobile health technology and his experience in clinical research, experiemention and digital innovation make him uniquely qualified to guide our ethical strategy. At Mindwell, Dr. Insel ensures our training processes align with the highest evidence-based psychiatric standards, advises us on privacy and data ethics and helps strengthen our credibility with regulators, providers and corporate clients through overseeing our ethical safeguards. Dr. Insel is an invaluable expert helping us bridging innovation and care.

2. Dr. Latanya Sweeney – Professor of Government and Technology at Harvard University

Dr. Latanya Sweeney is a foremost computer scientist and one of the earliest academics to study data privacy, digital accountability, and algorithmic ethics. The impact of Dr Sweeney´s research on re-identification risk catalyzed our modern social understanding of the broader implications regarding how anonymized data can still breach an individual's privacy. The impact of Dr. Sweeney's work is a critical concern to Mindwell as we optimize our chatbot with sensitive user data where complete anonymity is of the highest importance to our company.

Dr. Sweeney is a former Chief Technology Officer with the Federal Trade Commission and founder/director of Harvard's Public Interest Tech Lab. This extensive academic credibility, combined with decades of experience in data governance and the responsible development of technology, creates a foundation of knowledge that Dr. Sweeney brings. Her research will be especially useful for Mindwell's use of sensitive mental health data, ensuring transparency while guiding the organization in how it protects sensitive data as the privacy landscape evolves.

Dr. Latanya Sweeney has been central to Mindwell's development of our Bias Mitigation Task Force, assisting our company in safely handling sensitive mental health data. She provides key guidance toward Mindwell's goal of ensuring privacy compliance, transparency, and user protections in a shifting regulatory environment. Critical to our

company's structure of training data standards, ethical audits, and fairness benchmarks, Dr. Sweeney is an invaluable asset. Her continued oversight helps Mindwell uphold our strong commitment to ethical data [15].

3. Joy Buolamwini – Founder, Algorithmic Justice League

Joy Buolamwini is an leading advocate for algorithmic fairness and researcher at MIT Media Lab, with a significant focus on studying racial and gender bias particularly in facial recognition systems. Joy Buolamwini has a successful track record of arguing for significant infrastructure on AI development and public accountability to build awareness about racial and gender bias. Joy Buolamwini contributes expertise on fairness in AI and equity to Mindwell's intent regarding seeking to understand the intention of being just in the establishment of a digital mental health service [16].

Joy Buolamwini, is a leading voice in AI fairness, bringing vital expertise in equity and accountability to our company's ethical framework. As a member of Mindwell's ethics board, Jouy Buolamwini advises on fairness and inclusivity in LLM training and deployment, ensuring Mindwell's chatbot equitably understands and serves users from all racial, cultural, and linguistic backgrounds, helps Mindwell implement transparent audit mechanisms that monitor for bias in provider matching and user conversations and supports equity-first development, guiding our company's approach to ethical AI that represents all populations. Joy's leadership and critical insight into algorithmic fairness position her as an essential contributor to Mindwell's goal of building an inclusive, ethical, and effective AI-powered mental health platform.

4. YouTube Presentation
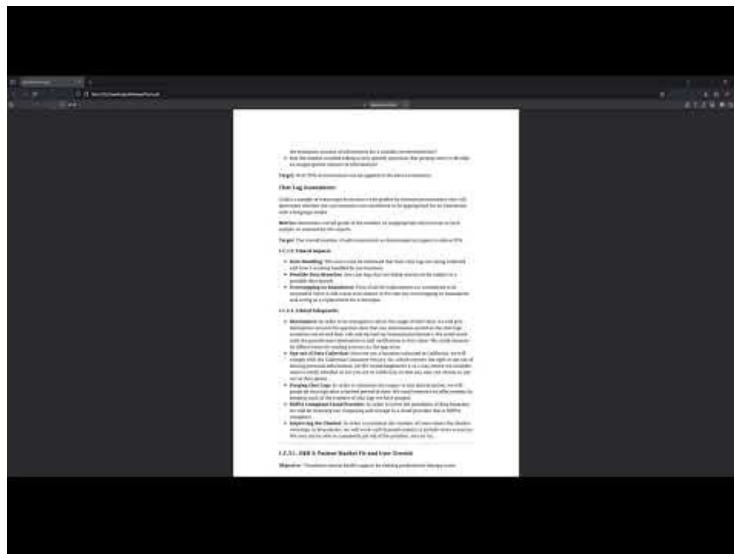
5. References

OKR 1 References:

1. J. I. Park et al., "Building Trust in Mental Health Chatbots: Safety Metrics and LLM-Based Evaluation Tools," arXiv.org, 2024. https://arxiv.org/abs/2408.04650 (accessed Mar. 09, 2025).

2. Interaction Design Foundation - IxDF, "What is Usability Testing?," Interaction Design Foundation - IxDF, Jun. 02, 2016. https://www.interaction-design.org/literature/topics/usability-testing (accessed Mar. 08, 2025).

3. "Differential Privacy and Applications - IEEE Digital Privacy," digitalprivacy.ieee.org. https://digitalprivacy.ieee.org/publications/topics/differential-privacy-and-applications (accessed Mar. 08, 2025).

OKR 2 References:

4. State of California Department of Justice, "California Consumer Privacy Act (CCPA)," State of California - Department of Justice - Office of the Attorney General, https://oag.ca.gov/privacy/ccpa (accessed Mar. 11, 2025).

OKR 3 References:

5. S. McCallum, "Meta settles Cambridge Analytica scandal case for $725m," BBC News, Dec. 23, 2022. [Online]. Available: https://www.bbc.com/news/technology-64075067. (Accessed: Mar. 10, 2025).

6. P. Voigt and A. V. D. Bussche, The EU General Data Protection Regulation (GDPR): Apractical guide. Springer International Publishing, 2017.

OKR 4 References:

7. "THOMAS LIU V. UBER TECHNOLOGIES, INC., No. 22-16507 (9th Cir. 2024)," Justia Law, 2024. [Online]. Available: https://law.justia.com/cases/federal/appellate-courts/ca9/22-16507/22-16507-2024-06-24.html.

Ethics Policy

8. "HIPAA Compliance - Amazon Web Services (AWS)," Amazon Web Services, Inc. https://aws.amazon.com/compliance/hipaa-compliance/] (https://aws.amazon.com/compliance/hipaa-compliance/ (accessed Mar. 11, 2025).

9. D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané, "Concrete Problems in AI Safety," arXiv preprint arXiv:1606.06565, Jun. 2016. [Online]. Available: https://arxiv.org/abs/1606.06565.

10. A. Abd-Alrazaq, Z. Safi, M. Alajlani, J. Warren, M. Househ, and K. Denecke, "Technical Metrics Used to Evaluate Health Care Chatbots: Scoping Review," Journal of Medical Internet Research, vol. 22, no. 6, p. e18301, Jun. 2020. [Online]. Available: https://doi.org/10.2196/18301.

11. Google Cloud, "GDPR Resource Center," Google, [Online]. Available:

https://cloud.google.com/security/compliance/gdpr. [Accessed: Apr. 10, 2025].

12. National Institute of Standards and Technology, "AI Risk Management Framework," NIST, [Online]. Available: https://www.nist.gov/itl/ai-risk-management-framework. [Accessed: Apr. 10, 2025].

13. Open Worldwide Application Security Project (OWASP), "OWASP Web Security Testing Guide," OWASP, [Online]. Available: https://owasp.org/www-project-web-security-testing-guide/. [Accessed: Apr. 10, 2025].

Bio References

14. Thomas Insel, M.D. "About - Thomas Insel, MD." [Online]. Available: https://www.thomasinselmd.com/about. [Accessed: Apr. 10, 2025].

15. L. Sweeney, "Dr. Latanya Sweeney's Home Page," latanyasweeney.org. [Online]. Available: https://latanyasweeney.org/indexlong.html. [Accessed: Apr. 10, 2025].

16. J. Buolamwini, "Overview," MIT Media Lab. [Online]. Available: https://www.media.mit.edu/people/joyab/overview/. [Accessed: Apr. 10, 2025].