An Introduction to AI for Health

Professor Charles Rahal

Leverhulme Centre for Demographic Science and Nuffield College

University of Oxford, 7th August, 2025

Introduction

Nice to meet you; I'm Charlie! I am a health data scientist!

- Associate Professor in Data Science & Informatics.
- Taught across the health and social sciences.
- Publish many things across these fields.
- Work presented at leading economics conferences.
- Specifically interested in methods development.
- A big part of this is AI for Healthcare

What will we learn today?

1. **Foundations of AI** Definitions, key concepts, and types of AI
2. **AI in Healthcare** From diagnostics to drug discovery
3. **Ethical & Regulatory Considerations** Privacy, bias, trust, and compliance
4. **Future Directions** Emerging technologies and career opportunities
5. **Getting Started** Resources, skillsets, and pathways to explore

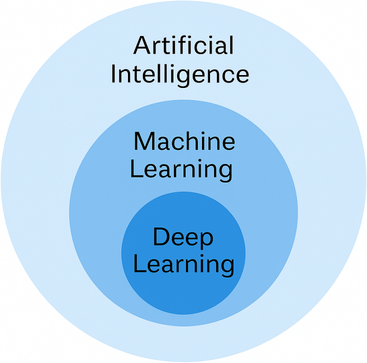
What is Artificial Intelligence (AI)?

- **Definition:** AI is the field concerned with the theory and development of computational systems that perform tasks that typically require human intelligence.
- **Core Objective:** Construct agents that perceive their environment and take actions that maximize expected utility.
- **Key Subfields:**
  - Machine Learning: Models that learn patterns from data.
  - Natural Language Processing: Understanding and generating human language.
  - Computer Vision: Extracting structure and meaning from visual inputs.
  - Planning and Reasoning: Algorithms for logic, search, and decision-making.

What is Artificial Intelligence (AI)? (continued)

- **Types of AI:**
    - Narrow AI: Domain-specific systems that perform single tasks, such as object recognition or translation.
    - General AI: Hypothetical agents with general reasoning and learning capabilities across multiple domains.

AI, Machine Learning, and Deep Learning



- **Artificial Intelligence (AI)** is the broadest field, encompassing all systems that mimic cognitive functions.
- **Machine Learning (ML)** is a subfield of AI that focuses on systems that learn from data without explicit programming.
- **Deep Learning (DL)** is a subset of ML that uses deep neural networks to model complex patterns in large datasets.

Relatable Examples of AI in Everyday Life

- **Voice Assistants (Siri, Alexa)**
  $\rightarrow$ Understand and respond to your questions
- **Google Search**
  $\rightarrow$ Ranks and suggests results using AI
- **Netflix / YouTube Recommendations**
  $\rightarrow$ Suggests content you're likely to watch
- **# Social Media Feeds (TikTok, Instagram)**
  $\rightarrow$ Algorithms decide what you see next
- **Google Maps / Waze**
  $\rightarrow$ Predicts traffic and best routes

How Do *You* Use AI?

- Can you think of a time you used AI today?
- What's the first app you open in the morning?
- Have you ever had a playlist or video 'just get you'?
- Have you spoken to a chatbot recently?

Bonus: Can you think of any other ways AI might impact your life that aren't obvious?

How Do *You* Use AI?

- Can you think of a time you used AI today?
- What's the first app you open in the morning?
- Have you ever had a playlist or video 'just get you'?
- Have you spoken to a chatbot recently?

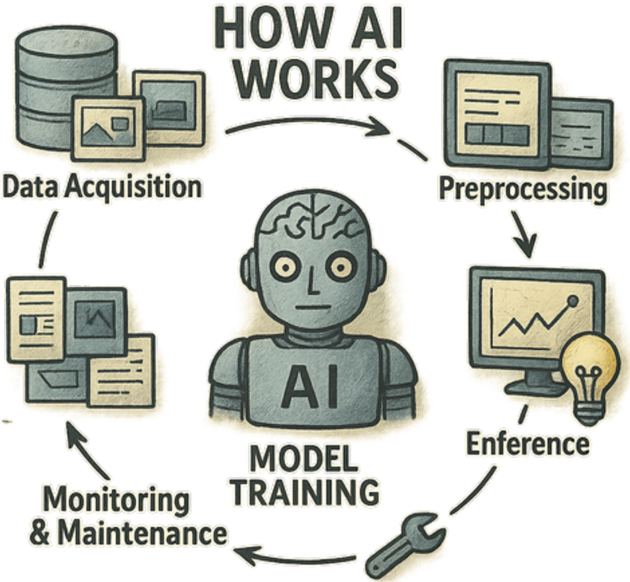
Bonus: Can you think of any other ways AI might impact your life that aren't obvious?

If you use a phone, you're using AI!


EDUCATION
• AI Tutors
• Essay/Grammar Helpers
• Auto-Generated Summaries
HEALTH
• Wearables
• Mental Health Apps
• Fitness Recommendations
ENTERTAINMENT
• Game NPCs
• Game Difficulty Scaling
• Music Recommen-
SHOPPING
• Smart Suggestions
• Virtual Try-On
• Price Prediction Tools
NEWS
NEWS
• AI-Filtered News Feeds
• Deepfake Defection Tools
• Content Moderation

Overview of AI Workflow

- **Data Acquisition:** Collection of raw data from various sources.
- **Preprocessing:** Cleaning, normalization, feature extraction.
- **Model Training:** Optimization of parameters on training set.
- **Evaluation:** Assessment on validation/test data; metrics.
- **Inference:** Deployment for predictions on new inputs.
- **Monitoring & Maintenance:** Performance tracking and retraining.

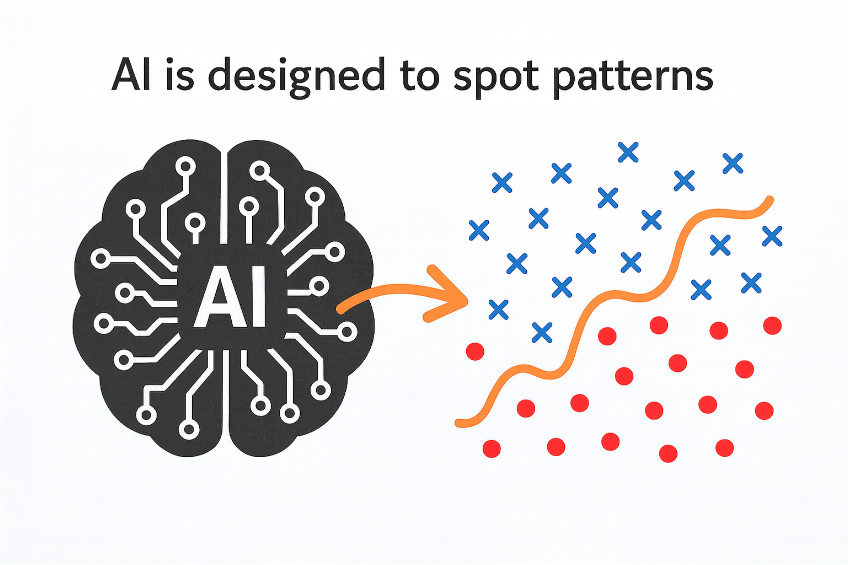

HOW AI
WORKS
Data Acquisition
Preprocessing
Enference
MODEL
TRAINING
AI
Monitoring
& Maintenance

AI is Pattern Recognition

- AI systems learn from data by identifying statistical or structural patterns.
- These patterns are used to map inputs (e.g., images, text, signals) to outputs (e.g., labels, actions).
- Training involves fitting a model to past observations.
- Success depends on how well learned patterns generalize to new data.
- There is no inherent “understanding” — only pattern exploitation.

Key Message

AI does not reason; it interpolates patterns found in historical data.


AI is designed to spot patterns
AI

Why Health? Why AI?

Importance of Healthcare

- Essential to societal welfare and productivity
- Increasing global burden of chronic and acute diseases
- Rising costs: health expenditure exceeding 10% of GDP in many nations

Challenges in Healthcare Systems

- Disease complexity and comorbidity
- Economic unsustainability
- Bottlenecks in diagnosis and treatment efficiency

Why AI?

- Scalable analysis of high-dimensional health data
- Predictive models for early detection and risk stratification
- Automation of diagnostics and administrative workflows

Simplified Example: Predicting Diabetes (1/2)

Problem Setup

- **Task:** Binary classification — predict onset of diabetes
- **Input:** Feature vector $x \in \mathbb{R}^d$
- **Output:** Label $y \in \{0, 1\}$
- **Goal:** Learn a function $f : \mathbb{R}^d \to [0, 1]$, where

$$\hat{y} = \mathbb{I}[f(x) > \tau]$$

Data Example (Pima Indians Diabetes Dataset)

- $d = 8$ features: Glucose, BMI, Age, BloodPressure, Insulin, etc.
- $n \approx 768$ labelled patient records

Simplified Example: Predicting Diabetes (2/2)

Learning Approach

- Supervised learning with training set $\{(x^{(i)}, y^{(i)})\}_{i=1}^{n}$
- Algorithms: Logistic Regression, Decision Trees, Random Forests, etc.
- Evaluation: Accuracy, Precision, Recall, ROC AUC

Model Interpretation

- Feature importance: Glucose and BMI typically have highest weight
- Decision threshold τ: Controls sensitivity vs specificity
- Outputs are probabilistic: $f(x) = \mathbb{P}(y = 1 \mid x)$

Types of Health Data Used in AI

- **Medical Images**
  MRI, X-ray, CT scans—AI analyzes pixels to detect patterns
- **Textual Records**
  Electronic health records, clinical notes, discharge summaries
- **Genomic Data**
  DNA sequences, gene expression profiles for precision medicine
- **Wearable Sensor Data**
  Heart rate, activity, sleep patterns from smartwatches and fitness trackers

AI Modalities: Data Types

- **Tabular**
  Risk scores, credit models
- **Text**
  Clinical notes, chatbots
- **Images**
  X-rays, MRIs
- **Time Series**
  ECG, wearables
- **Graphs**
  Protein networks, social graphs
- **Multimodal**
  Image and text, video captions

GPU vs CPU for AI

CPU

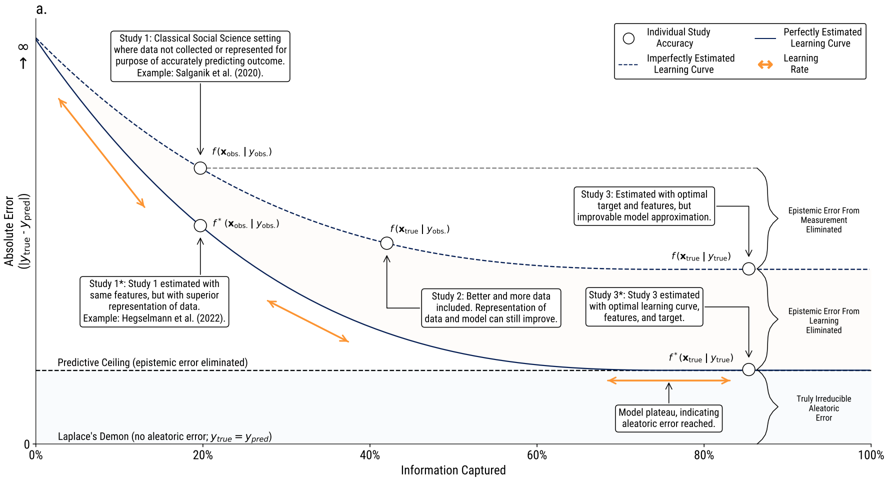
- Few, powerful cores
- Good for control flow / branching
- Lower peak throughput
- Large cache, complex logic
- Flexible, general-purpose

GPU

- Thousands of simple cores
- Excellent for parallelizing
- High memory bandwidth
- Optimized for dense linear algebra
- Training/inference at scale

All AI is actually about making predictions (of some kind).

But how accurate can predictions become in the future?


a.
Study 1: Classical Social Science setting where data not collected or represented for purpose of accurately predicting outcome. Example: Salganik et al. (2020).
Individual Study Accuracy
Perfectly Estimated Learning Curve
Imperfectly Estimated Learning Curve
Learning Rate
$f(\mathbf{x}_{obs.} \mid y_{obs.})$
$f^*(\mathbf{x}_{obs.} \mid y_{obs.})$
$f(\mathbf{x}_{true} \mid y_{obs.})$
Study 3: Estimated with optimal target and features, but improvable model approximation.
$f(\mathbf{x}_{true} \mid y_{true})$
Epistemic Error From Measurement Eliminated
Study 1*: Study 1 estimated with same features, but with superior representation of data. Example: Hegselmann et al. (2022).
Study 2: Better and more data included. Representation of data and model can still improve.
Study 3*: Study 3 estimated with optimal learning curve, features, and target.
Epistemic Error From Learning Eliminated
$f^*(\mathbf{x}_{true} \mid y_{true})$
Predictive Ceiling (epistemic error eliminated)
Model plateau, indicating aleatoric error reached.
Truly Irreducible Aleatoric Error
Laplace's Demon (no aleatoric error; $y_{true} = y_{pred}$)
Absolute Error ($|y_{true}$ - $y_{pred}|$) → ∞
0
0%
20%
40%
60%
80%
100%
Information Captured

Can AI Diagnose Disease Better Than a Human?

1. **Yes, in specific domains:**
   - Radiology (e.g. tumour detection in X-rays, MRIs)
   - Dermatology (e.g. skin-lesion classification)
2. **No, not universally:**
   - Complex, multi-factorial cases
   - Rare diseases with little training data
     - However, this is changing!
3. **Depends on:**
   - Quality & quantity of data
   - Model validation & regulatory approval
   - Human oversight & interpretability

What deos the future look like? Discuss in pairs for two minutes.

arXiv > cs > arXiv:2405.02957

Search... All fields Search

Help | Advanced Search

Computer Science > Artificial Intelligence

[Submitted on 5 May 2024 (v1), last revised 17 Jan 2025 (this version, v3)]

Agent Hospital: A Simulacrum of Hospital with Evolvable Medical Agents

Junkai Li, Yunghwei Lai, Weitao Li, Jingyi Ren, Meng Zhang, Xinhui Kang, Siyu Wang, Peng Li, Ya-Qin Zhang, Weizhi Ma, Yang Liu

The recent rapid development of large language models (LLMs) has sparked a new wave of technological revolution in medical artificial intelligence (AI). While LLMs are designed to understand and generate text like a human, autonomous agents that utilize LLMs as their "brain" have exhibited capabilities beyond text processing such as planning, reflection, and using tools by enabling their "bodies" to interact with the environment. We introduce a simulacrum of hospital called Agent Hospital that simulates the entire process of treating illness, in which all patients, nurses, and doctors are LLM-powered autonomous agents. Within the simulacrum, doctor agents are able to evolve by treating a large number of patient agents without the need to label training data manually. After treating tens of thousands of patient agents in the simulacrum (human doctors may take several years in the real world), the evolved doctor agents outperform state-of-the-art medical agent methods on the MedQA benchmark comprising US Medical Licensing Examination (USMLE) test questions. Our methods of simulacrum construction and agent evolution have the potential in benefiting a broad range of applications beyond medical AI.

Subjects: **Artificial Intelligence (cs.AI)**
Cite as: arXiv:2405.02957 **[cs.AI]**
(or arXiv:2405.02957v3 **[cs.AI]** for this version)
https://doi.org/10.48550/arXiv.2405.02957

Access Paper:

- View PDF
- HTML (experimental)
- TeX Source
- Other Formats

BY-NC-SA view license

Current browse context:
cs.AI
< prev | next >
new | recent | 2024-05
Change to browse by:
cs

References & Citations
- NASA ADS
- Google Scholar
- Semantic Scholar

Export BibTeX Citation

Bookmark

AgentHospital (arXiv:2405.02957, Li et al., 2025)

- **Simulacrum:** A full "virtual hospital" where patients, nurses, and doctors are all LLM-powered agents.
- **Agent Evolution:** Doctor agents train by treating tens of thousands of simulated patients—no manual labels needed.
- **State & Actions:** States = patient vitals/labs over time; Actions = diagnostic tests & treatment decisions.
- **Reward Signal:** Composite metric combining survival likelihood and recovery speed.
- **Performance:** Evolved agents outperform state-of-the-art medical-agent baselines by a significant margin.
- **Implications & Next Steps:** Potential for accelerated agent design in other domains. Future: integrate real-world EHR data/human–agent collaboration

Case Study: Google Flu Trends Overestimation

- **Background:** Launched 2008 to estimate influenza-like illness (ILI) rates from search query volumes
- **Initial Success:** Closely matched CDC ILI data through early 2012
- **Failure Mode:**
  - Overestimated 2013–14 U.S. flu season by up to 50
  - Predicted an early "second wave" that never occurred
- **Root Causes:**
  - Algorithmic overfitting to seasonal search spikes
  - Media-driven search behavior not tied to actual illness
  - Lack of transparency and model updates
- **Lessons Learned:**
  - Combine big-data models with traditional surveillance
  - Regular retraining and validation against ground truth
  - Importance of model interpretability in public health

AI in Diagnostics: Medical Imaging

AI-driven analysis enhances diagnostic accuracy and efficiency:

- **Core Techniques:**
  - Convolutional Neural Networks (CNNs) for feature extraction
  - U-Net architectures for image segmentation
  - Transfer learning to leverage pre-trained models
- **Key Applications:**
  - X-ray: Detection of pneumonia, fractures, lung nodules
  - MRI: Segmentation of brain tumors, multiple sclerosis lesions
  - CT: Automated quantification of emphysema/detection embolism
- **Advantages:**
  - Improved sensitivity & specificity over manual reads
  - Rapid triage and prioritization of critical cases
  - Quantitative measures (volume, density) for treatment monitoring
- **Example Study:**
  - CNN achieving radiologist-level pneumonia detection (Rajpurkar et al., 2017)

AI and Personalized Medicine

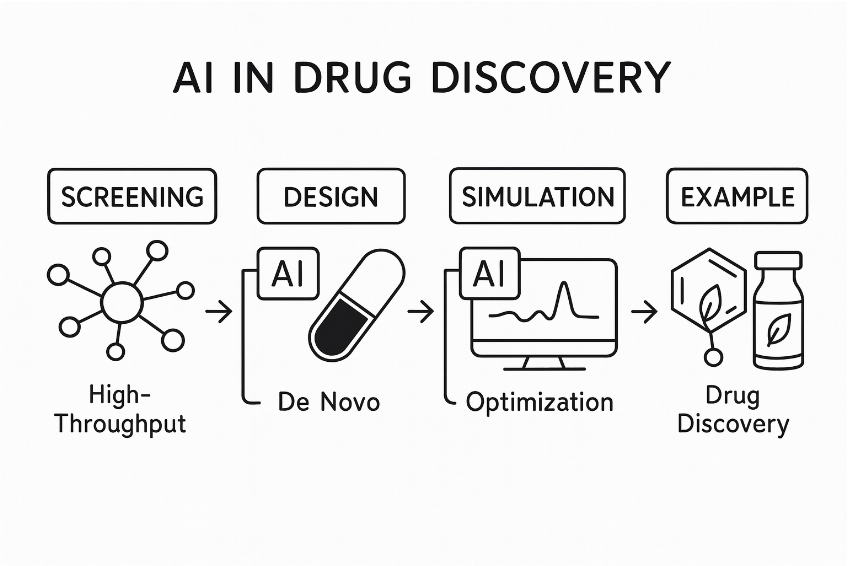
AI enables treatments tailored to each patient's genetic profile:

- **Genomic Profiling:** Machine learning models analyze DNA variants, gene expression, and biomarkers
- **Predictive Modeling:** AI forecasting of drug response and adverse effects based on genomic data
- **Treatment Optimization:** Selection of therapies (e.g., targeted drugs, immunotherapies) most likely to succeed
- **Example – Precision Oncology:**
  - Tumor sequencing identifies actionable mutations (e.g., EGFR, ALK)
  - AI recommends kinase inhibitors or checkpoint inhibitors
  - Demonstrated improved response rates in lung and breast cancers

AI and Drug Discovery

AI dramatically accelerates the development of new therapeutics:

- **High-Throughput Screening:** ML models predict molecule–target interactions, narrowing candidates
- **De Novo Molecule Design:** Generative models propose novel compounds with optimal properties
- **Optimization & Simulation:** AI guides in silico assays of efficacy, toxicity, and pharmacokinetics
- **Example – COVID-19 Vaccines:**
  - ML-driven antigen selection and epitope prediction
  - Accelerated mRNA vaccine candidate design and preclinical validation
  - Reduced development timeline from years to months


AI IN DRUG DISCOVERY
SCREENING
DESIGN
SIMULATION
EXAMPLE
AI
AI
High-Throughput
De Novo
Optimization
Drug Discovery

Protein Folding + AI = Big Discovery!

What's the Problem?

- Proteins are chains of building blocks (amino acids).
- To work properly, they need to fold into just the right 3D shape.
- Figuring out that shape is really hard!

What Can AI Do?

- AI looks at thousands of known protein shapes.
- It learns patterns and rules about how they fold.
- Then it can predict the shape of new proteins — fast and accurately!

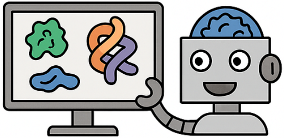
Why It Matters:

- Helps scientists understand diseases.
- Speeds up new medicine discovery.
- Opens doors in biotechnology and healthcare.

Protein Folding + AI = Big Discovery!

What's the Problem?

- Proteins are chains of building blocks (amino acids)
- To work properly, they need to fold into just the right 3D shape
- Figuring out that shape is really hard!

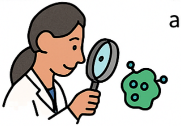


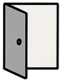
What Can AI Do?

- AI looks at thousands of known protein shapes
- It learns patterns and rules about how they fold — fast and accurately!

Why It Matters:

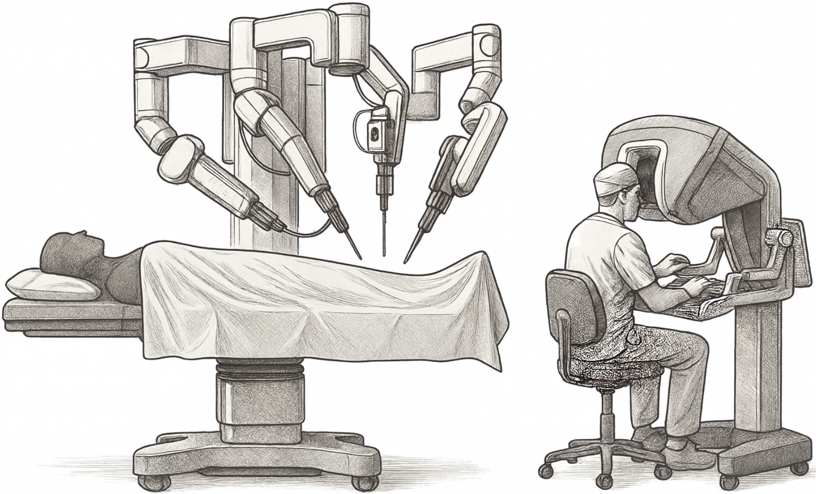
- Helps scientists understand diseases





Robotics and AI in Surgery

- **Robotic Surgical Systems**
  - Da Vinci® Surgical System: tele-operated robotic arms
  - Versius®: modular bedside units for minimally invasive procedures
- **AI-Enhanced Capabilities**
  - Real-time motion scaling and tremor suppression
  - Automatic recognition of anatomical landmarks
  - Augmented reality overlays to guide margin assessment
- **Clinical Impact**
  - Smaller incisions, reduced blood loss
  - Shorter hospital stays and faster postoperative recovery
  - Greater precision in complex operations

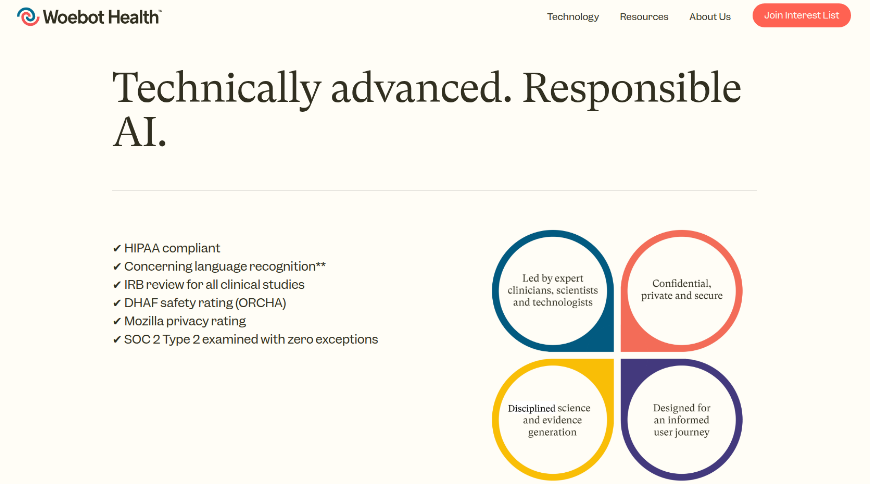


AI for Mental Health

AI is transforming mental health care through conversational agents and early-warning systems:

- **Mental Health Chatbots**
  - Conversational agents (e.g., Woebot); LLMs provide CBT-based support.
  - 24/7 accessibility, stigma-free interaction, mood tracking over time.
- **Predictive Models for Crises**
  - Machine learning on social media text, digital phenotyping.
  - Early identification of suicide risk, major depressive episodes.
  - Enables timely intervention by clinicians or emergency services.

Discussion: What are the ethical considerations around privacy and accuracy?


Woebot Health™
Technology
Resources
About Us
Join Interest List
Technically advanced. Responsible AI.
✓ HIPAA compliant
✓ Concerning language recognition**
✓ IRB review for all clinical studies
✓ DHAF safety rating (ORCHA)
✓ Mozilla privacy rating
✓ SOC 2 Type 2 examined with zero exceptions
Led by expert clinicians, scientists and technologists
Confidential, private and secure
Disciplined science and evidence generation
Designed for an informed user journey

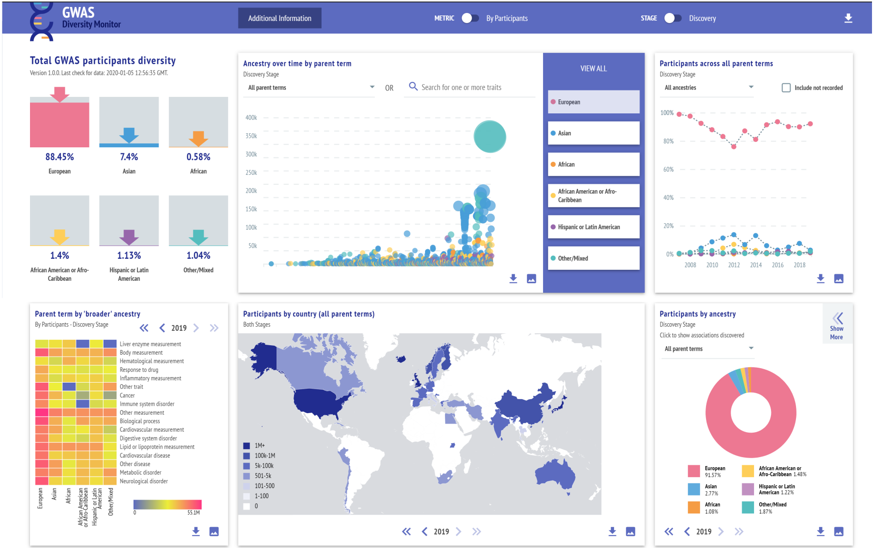
Ethical Questions in AI for Healthcare

- **Patient Privacy**
  - Collection and storage of sensitive health data
  - Risk of re-identification in "anonymized" datasets
  - Informed consent for secondary use of patient records
- **Data Security**
  - Safeguarding data at rest and in transit (encryption, access control)
  - Vulnerabilities in cloud-based AI platforms
  - Potential for malicious attacks (data poisoning, model inversion)
- **Regulatory Compliance**
  - Adherence to GDPR, HIPAA, and emerging AI-specific regulations
  - Auditability and transparent reporting of AI decision processes

Bias in AI Systems

- **Definition of Bias** Systematic error arising when training data or model assumptions do not represent the target population.
- **Mechanisms of Harm**
  - Under-representation: Minority groups poorly represented
  - Proxy variables: Non-clinical attributes (e.g. ZIP code) inadvertently encode socio-economic status
  - Feedback loops: Biased predictions reinforce inequitable data collection
- **Real-World Example** Health-risk prediction algorithm (Obermeyer et al., 2019): Underestimated risk for Black patients; delayed referrals
- **Mitigation Strategies**
  - Audit models on stratified subgroups
  - Re-balance training data or apply fairness constraints
  - Incorporate domain expertise to adjust cost functions

Bias in AI


GWAS
Diversity Monitor
Additional Information
METRIC By Participants
STAGE Discovery
Total GWAS participants diversity
Version 1.0.0. Last check for data: 2020-01-05 12:56:35 GMT.
88.45% European
7.4% Asian
0.58% African
1.4% African American or Afro-Caribbean
1.13% Hispanic or Latin American
1.04% Other/Mixed
Ancestry over time by parent term
Discovery Stage
All parent terms
OR
Search for one or more traits
VIEW ALL
European
Asian
African
African American or Afro-Caribbean
Hispanic or Latin American
Other/Mixed
Participants across all parent terms
Discovery Stage
All ancestries
Include not recorded
Parent term by 'broader' ancestry
By Participants - Discovery Stage
2019
Liver enzyme measurement
Body measurement
Hematological measurement
Response to drug
Inflammatory measurement
Other trait
Cancer
Immune system disorder
Other measurement
Biological process
Cardiovascular measurement
Digestive system disorder
Lipid or lipoprotein measurement
Cardiovascular disease
Other disease
Metabolic disorder
Neurological disorder
Participants by country (all parent terms)
Both Stages
1M+
100k-1M
5k-100k
501-5k
101-500
1-100
0
Participants by ancestry
Discovery Stage
Click to show associations discovered
All parent terms
Show More
European 91.57%
Asian 2.77%
African 1.08%
African American or Afro-Caribbean 1.48%
Hispanic or Latin American 1.22%
Other/Mixed 1.87%

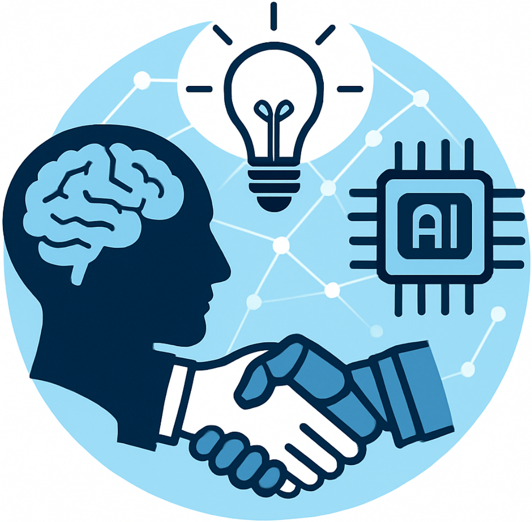
Reliability and Trust

- **Types of AI Errors**
  - False positives: incorrect flagging of disease; unnecessary interventions
  - False negatives: missed diagnosis leads to delayed treatment
  - Calibration errors: probability estimates misaligned with true risk
- **Consequences of Mistakes**
  - Patient harm: physical, psychological, financial
  - Erosion of clinician and patient trust in AI tools
  - Legal and regulatory liability for developers and providers
- **Mitigation Strategies**
  - Rigorous validation on diverse, representative datasets
  - Continuous monitoring and post-market surveillance
  - Human–AI collaboration: "human in the loop" for ambiguous cases
- **Building Trust**
  - Explainable AI: transparent decision processes and confidence scores
  - Clinician training: understanding AI strengths and limitations
  - Clear protocols for escalation when AI outputs conflict with clinical judgment

Would you trust an AI to diagnose your health?

- **Discuss for 2 minutes:**
  - What factors would make you trust an AI diagnosis?
  - What concerns or risks would make you hesitate?
  - When would you trust AI more or less?
- **Be ready to share:**
  - One key argument for trusting AI.
  - One key argument against trusting AI.

We require strong cooperation between AI and humans

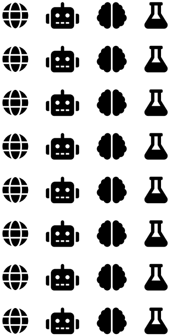


Regulation and AI in Healthcare

- **Why Regulate?**
  - Ensure patient safety and efficacy of AI tools
  - Protect privacy and data rights
  - Prevent misuse and malicious applications
- **Key Regulatory Domains**
  - Validation & Approval: clinical trials, FDA/EMA clearance
  - Data Governance: GDPR, HIPAA, data provenance
  - Transparency: model documentation (e.g. Datasheets for Datasets)
  - Liability & Accountability: defining responsibility for errors

Future of AI in Health

- **Digital Twins**
  - Virtual replicas of patients for in silico trials
- **Federated Learning**
  - Collaborative model training without data sharing
- **Explainable AI (XAI)**
  - Transparent models for clinician trust
- **AI-Driven Drug Design**
  - Generative models discovering novel compounds



Exciting Future Possibilities

- **Personalized Health Assistants**
  - Proactive coaching via wearables and chatbots
- **Nanotechnology + AI**
  - Targeted drug delivery & real-time diagnostics
- **Brain–Computer Interfaces**
  - Restoring mobility & communication for patients
- **AI-Powered Epidemiology**
  - Early outbreak detection via social media & mobility data

AI Health Careers

- **Data Scientist**
  - Build & validate predictive models on clinical data
- **Biomedical Engineer**
  - Design AI-enabled medical devices & imaging systems
- **Health AI Researcher**
  - Advance algorithms in genomics, radiology, and EHR analysis
- **Clinical AI Specialist**
  - Integrate & monitor AI tools in hospital workflows

Skillset Needed

- **Mathematics** Linear algebra, probability, statistics
- **Computer Science** Algorithms, data structures, software engineering
- **Biology & Medicine** Physiology, pathology, clinical protocols
- **Ethics & Policy** Privacy, fairness, regulatory frameworks
- **Problem-Solving** Critical thinking, interdisciplinary collaboration
- **Communication** Explaining technical concepts to clinicians and patients

How to Get Started in AI

- **Academic Pathways**
  - A-level: Mathematics, Computer Science, Biology, Physics
- **Online Courses**
  - Coursera: Andrew Ng's Machine Learning; fast.ai; edX MicroMasters
- **University Clubs & Hackathons**
  - AI societies; Kaggle competitions; biomedical datathons
- **Open-Source Projects**
  - Contribute to TensorFlow, PyTorch, MONAI, scikit-learn
- **Internships & Mentorship**
  - Seek placements in hospitals, startups, research labs

Conclusion

- AI is revolutionizing healthcare, from diagnostics to drug discovery.
- Ethical considerations are crucial for responsible AI deployment.
- The future holds exciting possibilities, but challenges remain.
- Requires interdisciplinary skills/commitment to improve patient outcomes.

Conclusion

- AI is revolutionizing healthcare, from diagnostics to drug discovery.
- Ethical considerations are crucial for responsible AI deployment.
- The future holds exciting possibilities, but challenges remain.
- Requires interdisciplinary skills/commitment to improve patient outcomes.

Thank you for your attention! Happy to answer any questions!