

Uses of AI

BENEFITS AND LIMITATIONS

J. Langley,
Founder, Huntsville AI
CTO, CohesionForce, Inc
January 16, 2025

INTRODUCTION TO J. LANGLEY

Lots of titles:

- Chief Technical Officer
- Chief Archeologist
- Chief Instigator

I've been working with AI since around 2005, with my master's project at Florida Institute of Technology being an NLP based system for recommending web forum channels that best match a user's question.

I believe that the best way to ensure that AI is used for the greater good is to involve the greatest number of perspectives possible in its development, testing, and use.

AI TECHNIQUES

1. Natural Language Processing - Classification, Summarization, Generation
2. Computer Vision - Classification, Segmentation, Enhancement, Generation
3. Audio Processing - Recognition, Classification, Generation
4. Multimodal Tasks - Text-to-Image, Text-to-Sound, Text-to-Video
5. Reinforcement Learning - Autonomous Vehicles, Robotics, Optimization
6. Recommendation Systems - Collaborative Filtering, Content Based, Implicit
7. Generative AI - Text, Images, Audio, Video
8. Agentic AI - Tool Calling, Decision Making
9. Robotics - Humanoid Robots, Autonomous Drones, Self Driving Cars

NATURAL LANGUAGE PROCESSING

- Classification
- Named Entity Recognition
- Sentiment Analysis
- Summarization
- Semantic Understanding
- Language Translation

NATURAL LANGUAGE PROCESSING

Datasets

- Penn Treebank
- Wikipedia: A vast repository of encyclopedic text, often used for pretraining language models.
- Common Crawl: A publicly available dataset of web pages, used for large-scale training.
- Penn Treebank (PTB): A widely used dataset for syntactic parsing and language modeling.
- Yelp Reviews: Used for sentiment analysis tasks.
- CNN/DailyMail: Articles paired with their summaries.
- GLUE (General Language Understanding Evaluation): A suite of tasks for evaluating NLP models.

NATURAL LANGUAGE PROCESSING

Limitations

- Language Bias
- Gender Bias
- Racial / Ethnic Bias
- Detecting sarcasm or dark humor
- Words change meaning over time
- Implied context
- Model context size
- Emerging language (generational problems)

COMPUTER VISION

- Classification
- Segmentation
- Enhancement
- Object Recognition
- Facial Recognition
- Pose Recognition

COMPUTER VISION

Datasets

- ImageNet: A large-scale dataset with over 14 million labeled images across 1,000 classes; foundational for many state-of-the-art models.
- MNIST: A dataset of 70,000 grayscale images of handwritten digits (0–9); widely used for entry-level tasks.
- COCO (Common Objects in Context): A large-scale dataset with over 330,000 images, including annotations for object detection, segmentation, and keypoints.
- VGGFace2: A dataset with over 3 million images for face recognition.
- Human3.6M: A large-scale dataset for 3D human pose estimation
- LUNA16: A dataset for lung nodule detection in CT scans.
- Waymo Open Dataset: A dataset for autonomous vehicle perception, including lidar and camera data.

COMPUTER VISION

Limitations

- Gender Bias
- Racial / Ethnic Bias
- Images without context
- Difficulty translating to real world implementation
- Adversarial vulnerabilities by changing small parts of images
- Lack of physical world understanding

AUDIO PROCESSING

- Speech Recognition
- Speaker Recognition
- Sentiment Analysis
- Classification
- Source Separation
- Speaker Diarization

AUDIO PROCESSING

Datasets

- LibriSpeech: Audiobooks with transcriptions, often used for training ASR models.
- Common Voice (Mozilla): A large dataset of diverse voices, crowdsourced for ASR.
- MUSDB18: A dataset for audio source separation, such as isolating vocals and instruments.
- IEMOCAP: Speech recordings annotated with emotion labels.
- CREMA-D: A dataset of acted emotional speech with diverse speakers.
- Macaulay Library (Cornell Lab of Ornithology): A massive archive of bird and other animal sounds.

AUDIO PROCESSING

Limitations

- Limited training data diversity and domains
- Regional accents, phrases, 'code switching'
- Lack of semantic understanding (multiple English words sound alike)
- Lack of visual cues used by humans
- Have trouble with background noise or other effects

MULTIMODAL TASKS

Image-to-Text

- Generating descriptive textual captions for a given image.
- Answering questions based on the content of an image.
- Retrieving the most relevant image given a textual query or vice versa.

Text-to-Image

- Generating images based on textual descriptions.

Audio-to-Text

- Generating textual descriptions of audio events.
- Speech Recognition

Text-to-Audio

- Generating speech from textual input.
- Generating music from textual input.

Audio/Video-to-Text

- Improved speech recognition by combining video and audio

MULTIMODAL TASKS

Datasets

- COCO (Common Objects in Context): Images with object annotations and captions.
- MSR-VTT: 10,000 video clips with textual captions.
- AVE (Audio-Visual Event Dataset): Video clips with audio-visual event labels.
- LRS3 (Lip Reading Sentences): Video clips with synchronized audio and transcriptions.
- CMU-MOSEI: Videos with aligned audio, text, and sentiment/emotion labels.

MULTIMODAL TASKS

Limitations

- Data Alignment (time alignment), scarcity for some domains, data diversity.
- Domain transfer problems (movie subtitles vs instructional videos)
- Interpretability or Explainability with troubleshooting problems
- Difficult to evaluate correctness - lack of benchmarks and metrics
- Imbalanced datasets (more text than audio than video)

RECOMMENDATION SYSTEMS

- Collaborative Filtering
- Content Based
- Implicit
- Personalized
- Context Aware

One of the earliest public uses of AI:

- Product placement in stores
- Rewards programs at checkout
- Google search results

RECOMMENDATION SYSTEMS

Datasets

- MovieLens 20M: 20 million ratings from 138,000 users.
- Netflix Prize Dataset: 100 million movie ratings from 480,000 users on 17,000 movies.
- Amazon Product Review Datasets: A large-scale dataset of user reviews and ratings from Amazon.
- Retailrocket Dataset: Data from an e-commerce platform. User events (clicks, purchases, add-to-cart).
- Expedia Dataset (Kaggle): User search data, booking details, and hotel information.
- Instacart Dataset (Kaggle): User purchase histories, product details, and order sequences.

RECOMMENDATION SYSTEMS

Limitations

- Can be creepy
- Cold start problem
- Data sparsity
- Bias:
 - Popularity Bias
 - Confirmation Bias
 - Demographic Bias
- Scalability based on number of products and users
- Privacy? What privacy?
- Dynamic content (recommending out of stock items)

GENERATIVE AI

- Generating Text - lots and lots of text.
 - Can be formatted in specific ways
 - Can match a given tone
- Images
 - Can create images from text descriptions.
 - Can alter images based on another provided image
- Audio
 - Can create audio from text
 - Can create music from text
- Video
 - Can create video from text
 - Short clips are fairly easy at the moment, with longer running videos coming soon

GENERATIVE AI

Datasets

- Keyed by advancements in large language models
- MAESTRO: Piano performances with aligned MIDI and audio files.
- LAION-5B: A massive dataset of image-text pairs scraped from the web.

GENERATIVE AI

Technical Limitations

- Plagiarism and Copyright Issues - Generated content may inadvertently copy from the training data, raising intellectual property concerns.
- Misinformation and Deepfakes - Generative AI can create highly convincing fake text, images, videos, and audio, making it easier to spread misinformation.
- Coherence in Long Outputs - Struggles to maintain coherence in long or complex narratives.

Cultural/Ethical Limitations

- Alignment to cultural norms
- Misuse by bad actors

ROBOTICS

- Humanoid Robots
- Autonomous Drones
- Collision Avoidance
- Manufacturing

ROBOTICS

Datasets

- NYU Depth Dataset: RGB-D (color and depth) images of indoor scenes.
- Waymo Open Dataset: Multi-sensor (Lidar, camera, radar) data for autonomous driving.
- DexNet (Dexterity Network): Synthetic dataset for robotic grasp planning, including object meshes and grasp quality metrics.
- OpenAI Gym: Standardized environments for testing reinforcement learning algorithms.
- Amazon Picking Challenge Dataset: Datasets from robotics challenges involving object picking in cluttered environments.

ROBOTICS

Limitations

- Human interaction - blinkers on robots, CMU Rover problem
- Driver manipulation of autonomous vehicles
- Reward Hacking for reinforcement learning

AGENTIC AI

- Tool Calling
- Decision Making - show your work
- Chain of Thought Reasoning

AGENTIC AI

Datasets

- I could not find datasets specific to this.

AGENTIC AI

Limitations

- Changing direction based in intermediate feedback is difficult
- Tools that interact with the real world are hard

Questions?

COMMENTS, OR CRIES OF HERESY?

J. Langley,
Founder, Huntsville AI
CTO, CohesionForce, Inc
January 16, 2025