AWS AI managed services
https://squasta.github.io
Updated: 10 06 2026

- **AWS Hardware for AI**
  - EC2
    - GPU based instances
      - P3-P5
      - G3-G6
    - AWS Trainium
      - ML Chips to perform Deep Learning
        - Trnl instances
      - Up to 50% lower training costs than comparable Amazon EC2 instances
      - Lowest environmental footprint
    - AWS Inferia
      - Chips for Deep Learning and generative AI inference
        - Inf, Inf2 instances
      - Up to 70% cost reduction
- **AI Service for medical**
  - Amazon Transcribe Medical
    - Automatically convert medical related speech to text
    - HIPAA Compliant
  - Amazon Comprehend Medical
    - NLP to detect Protected Health Information (PHI)
      - Realtime analyze
      - Analysis Jobs
  - AWS HealthScribe
    - An HIPAA-eligible service
    - To empower healthcare software vendors to build clinical applications
  - Amazon Connect Health
- **Amazon Augmented AI (A2I)**
  - To incorporate a human review into your workflow. Human oversight of ML predictions
    - Private (Company employees)
    - AWS contractors
    - AWS Mechanical Turkey
  - ML build on AWS or elsewhere
    - Ex: Amazon Sagemaker Human Review
- **Amazon Mechanical Turk**
  - A crowdsourcing marketplace that makes it easier for individuals and businesses to outsource their processes and jobs to a distributed workforce who can perform these tasks virtually
  - Access a global,on-demand, 24x7 workforce
  - Usage
    - Image classification
    - Data collection
- **Amazon Kendra**
  - A managed information retrieval and intelligent document search service
  - Uses natural language processing and advanced deep learning model
- **Amazon Textract**
  - To extract text, handwriting and data from scanned documents (OCR)
  - Analyze
    - Document
    - Expense
    - ID
- **AWS Personalize**
  - Fully managed service ML-apps to build apps with real-time personalised recommendations
  - Recipes
    - Amazon Algorithms prepared for specific use case
  - Usage
    - Retail stores
    - Media
- **ML Platform**
  - Amazon Sagemaker AI
    - Custom ML models
      - Build
      - Train
      - Deploy
- **GenAi**
  - Amazon Bedrock
  - Amazon Q
    - Amazon Q Business
    - Amazon Q Developer
- **Vision**
  - Amazon Rekognition
    - Image and Video Analysis
    - To find
      - Objects
      - People
        - Celebrity
      - Text
      - Scene
    - Usage
      - Labelling
      - Text detection
        - Custom labels
      - Face detection + analysis
      - Content moderation
        - Integrated with Amazon Augmented AI (A2I)
- **Text et NLP (Natural Langage Processing)**
  - Amazon Translate
    - Localize content for diverse global users and translate and analyze large volumes of text to activate cross-lingual communication between users
    - Real-time translation
    - Batch translation
      - Translate Jobs
    - Custom terminology
    - Parallel data
      - To influence your translation (style...)
  - Amazon Comprehend
    - Extract sentiment and insight from text
    - Custom Classification
    - Named Entity Recognition (NER)
      - Extracts predefined, general purpose entities from text
    - Custom Entity Recognition
- **Audio and Chat**
  - Amazon Transcribe
    - Speech to text
      - Using Deep Learning process named Automatic Speech Recognition (ASR)
    - Can remove PII during recognition
    - Custom vocabularies (for words)
      - Use both for the highest transcription accuracy
    - Custom language models (for context)
      - To improve transcription accuracy for domain-specific speech
    - Toxicity detection
  - Amazon Polly
    - Text to speech
    - Lexicons
      - To customize the pronunciation of words
    - Speech Synthesis Markup Language (SSML)
      - To customize and control aspects of speech such as pronunciation, volume, and speech rate
    - Voice Engine
    - Speech Mark
      - Metadata that describe the speech that you synthesize
  - Amazon Lex
    - To build Chatbot quickly
      - Text
      - Voice
    - Creation method
      - Traditional
      - GenAI
        - With AWS Bedrock