Name	Definition	Notes
Amazon Rekognition	Amazon Rekognition is a machine learning-based image and video analysis service that uses convolutional neural networks (CNNs) to detect objects, faces, text, and scenes. It supports facial recognition, facial analysis, activity recognition, content moderation, and real-time video processing through APIs, enabling scalable integration into applications for security, marketing, and automation.	Uses = Image classification, object detection, detection of text in image, facial recognition, sentiment, and public saftey
Amazon Textract	Amazon Textract is a fully managed machine learning service that uses deep learning models to extract text, forms, and tables from scanned documents. It applies Optical Character Recognition (OCR) in conjunction with Natural Language Processing (NLP) to analyze document layouts, identify key-value pairs, and detect structured data. Textract can process both printed and handwritten text, enabling automated document processing workflows with high accuracy and scalability.	
Amazon Transcribe	Amazon Transcribe is a fully managed automatic speech recognition (ASR) service that converts spoken language into text. It uses deep learning models to transcribe audio from various sources (e.g., calls, meetings) in real-time or batch mode, supporting multiple languages, speaker identification, and custom vocabulary for enhanced accuracy. The service also offers features like punctuation, timestamps, and sentiment analysis for more context-rich transcriptions.	
Amazon Translate	Amazon Translate is a neural machine translation (NMT) service that leverages advanced deep learning models, including transformer-based architectures, to provide high-quality automatic translation of text. It supports real-time and batch translation, enabling dynamic language pairings and context-aware translation. The service allows customization through user-defined terminology and domain-specific vocabularies, optimizing translation accuracy. Amazon Translate scales horizontally, integrating easily into applications and workflows with APIs for both structured and unstructured text processing.	
Amazon Polly	Amazon Polly is a neural text-to-speech (TTS) service that uses deep learning models to convert text into natural-sounding speech. It offers a wide selection of languages and voices, with features like real-time streaming, SSML support, and customizable speech parameters for integration into applications, devices, and services.	
Amazon Lex	Amazon Lex is a fully managed service for building conversational interfaces using natural language understanding (NLU) and automatic speech recognition (ASR). It enables the creation of chatbots and voice applications by processing text and speech inputs to understand user intent. Lex leverages deep learning models to handle context, slot management, and multi-turn conversations, with seamless integration into AWS services like Lambda, S3, and DynamoDB for enhanced functionality.	
Amazon Kendra	Amazon Kendra is an AI-driven enterprise search service that uses advanced natural language processing (NLP) and deep learning techniques to deliver contextually relevant search results. It indexes structured and unstructured data from multiple sources (e.g., SharePoint, Salesforce, S3, and databases) and applies query understanding, semantic search, and entity recognition to interpret user queries. Kendra uses a transformer-based architecture for automatic relevance tuning and customizable ranking models, enabling fine-grained control over search results. The service supports integration with AWS Lambda for custom data processing and offers APIs for embedding intelligent search capabilities into applications.	
Amazon Personalize	"Amazon Personalize is a fully managed machine learning service that enables real-time, personalized recommendations using advanced algorithms like collaborative filtering, deep learning-based models (e.g., neural networks), and factorization machines. It automatically preprocesses data (e.g., user activity, item interactions) and optimizes model training and hyperparameters, leveraging techniques such as implicit feedback modeling and personalized ranking. Personalize supports batch and real-time inference through APIs, with customizable recommendation strategies for content, product, and user engagement. It scales elastically and integrates with AWS services (e.g., S3, Lambda) to power personalized experiences across applications.	
Amazon Forecast	Amazon Forecast is a fully managed service for time-series forecasting that leverages state-of-the-art machine learning models, including DeepAR+, Prophet, and Fourier-based methods, to generate high-accuracy predictions. It automatically handles data preprocessing, feature engineering, and model selection, while enabling users to incorporate additional variables (e.g., weather, holidays) to improve forecast accuracy. Forecast uses a combination of deep learning (LSTM-based models) and classical statistical techniques to model complex temporal dependencies. The service supports both univariate and multivariate forecasting, providing real-time and batch inference through APIs. It integrates with AWS data sources (e.g., S3, Redshift) and scales elastically to handle large datasets and high throughput.	

Amazon Comprehend	Amazon Comprehend is a fully managed natural language processing (NLP) service that employs deep learning techniques, including transformer-based models (e.g., BERT and its variants), to perform a variety of text analysis tasks at scale. It offers built-in capabilities for entity recognition (e.g., PERSON, LOCATION, ORGANIZATION), sentiment analysis (fine-grained sentiment classification), key phrase extraction, language detection, and syntax analysis (tokenization, part-of-speech tagging). Comprehends entity recognition leverages Named Entity Recognition (NER) and contextual embedding methods for more accurate identification of entities within complex text.	
Amazon CodeGuru	Amazon CodeGuru is a fully managed service that uses machine learning and static code analysis to identify bugs, security vulnerabilities, and performance issues in source code. It consists of CodeGuru Reviewer, which analyzes pull requests to detect issues and suggest best practices (e.g., performance, security, maintainability), and CodeGuru Profiler, which analyzes application runtime to identify bottlenecks and optimize resource usage. CodeGuru supports Java, Python, and JavaScript, integrates with Git-based repositories (e.g., GitHub, Bitbucket, AWS CodeCommit), and can be embedded into CI/CD pipelines for continuous quality monitoring. The service leverages large-scale ML models trained on extensive codebases for context-aware recommendations.	
Amazon Augmented AI	Amazon Augmented AI (A2I) is a managed service that facilitates human-in-the-loop (HITL) workflows, enabling human review and oversight of machine learning predictions. It integrates with AWS AI services (e.g., Amazon Rekognition, Textract, Comprehend) and custom models to manage tasks that require human judgment, such as data validation, classification, or complex decision-making. A2I provides a framework for defining human review tasks, configuring approval workflows, and tracking reviewer decisions. It supports dynamic routing of data to human reviewers through integration with Amazon Mechanical Turk or private workforce solutions, allowing for scalable and secure review processes. A2I also offers built-in monitoring and analytics to ensure model accuracy and continuous performance improvement.	
Amazon SageMaker	Amazon SageMaker is a fully managed machine learning platform that provides a comprehensive suite of tools for the entire ML lifecycle. It includes SageMaker Studio for integrated development, offering Jupyter notebooks, debugging, and visualization capabilities. SageMaker supports distributed training and model tuning using Amazon SageMaker Training Jobs and Hyperparameter Optimization with algorithms like XGBoost, Linear Learner, and deep learning frameworks (e.g., TensorFlow, PyTorch, MXNet) via custom Docker containers. It also provides SageMaker Autopilot for automated model building and SageMaker Ground Truth for scalable data labeling with active learning. For model deployment, SageMaker offers managed endpoints for real-time inference and batch transform for large-scale inference jobs, while SageMaker Model Monitor tracks model drift and performance in production. SageMaker Neo enables model optimization for deployment on edge devices, converting models into highly efficient formats that run on cloud or edge hardware. SageMaker integrates seamlessly with AWS Data Wrangler, AWS Lambda, S3, and Redshift for end-to-end workflows, and supports versioned model management with SageMaker Model Registry. SageMaker Pipelines allows for CI/CD of ML models, automating workflows for data processing, training, validation, and deployment, enabling scalable, production-ready machine learning applications. The service also provides custom classification and entity recognition, enabling users to train domain-specific models using supervised learning on labeled datasets. These models are built using transfer learning and fine-tuning on task-specific data, allowing for high flexibility in identifying custom entities or categorizing text into user-defined classes. For topic modeling, Comprehend applies unsupervised learning techniques like Latent Dirichlet Allocation (LDA) and neural topic modeling to identify clusters of related documents or topics within large corpora.	