AI Data Management and Activation Cheat Sheet for Doerscircle

Foundation: Build the Basics

- **Data Governance**
  - Assign a Data Officer (formal or informal) to oversee data-related decisions.
  - Define clear data ownership and responsibility for each use case.
  - Establish guidelines for data quality, privacy, and compliance (e.g., GDPR).
- **Data Privacy & Security**
  - Implement robust data encryption (in transit and at rest).
  - Use anonymization techniques for customer data.
  - Regularly audit access controls and permissions.
  - Partner with trusted vendors for hosting and processing data securely.
- **Data Infrastructure**
  - Centralize data storage in a scalable cloud platform (e.g., Azure, AWS, or Google Cloud).
  - Use database solutions suited to your needs (structured: SQL databases, unstructured: NoSQL databases).
  - Develop APIs to ensure seamless data exchange across systems.

Step 1: Preparation for Use Cases

- **Simple ChatBot for Automated Customer Service**
  - Collect: Historical customer service logs, FAQs, and email/chat transcripts.
  - Clean: Remove sensitive data and standardize format.
  - Organize: Create a labeled dataset (queries and ideal responses).
  - Tools: Consider plug-and-play chatbot platforms like Microsoft Bot Framework or Intercom.
- **Customer Sentiment Analysis**
  - Collect: Customer feedback (emails, reviews, social media comments).
  - Clean: Remove duplicates and irrelevant entries (e.g., spam).
  - Organize: Tag data with sentiment labels (positive, negative, neutral) for training models.
  - Tools: Use tools like Hugging Face sentiment models or Google Cloud NLP.
- **Personalized Service Recommendations**
  - Collect: Customer behavior data (purchase history, browsing patterns, preferences).
  - Clean: Merge fragmented data points across channels.
  - Organize: Develop a schema for products, services, and customer attributes.
  - Tools: Utilize recommendation engines like Amazon Personalize or Google Recommendations AI.
- **Predictive Analytics for Customer Needs**
  - Collect: Historical sales and engagement data.
  - Clean: Identify and handle missing data.
  - Organize: Aggregate data into time-series formats.
  - Tools: Leverage platforms like Tableau or Microsoft Power BI with predictive modeling.
- **Simple Internal LLM Model**

- Collect: Internal documents (guidelines, policies, training material).
- Clean: Remove outdated or irrelevant documents.
- Organize: Structure content in clear categories for easier model training.
- Tools: Start with fine-tuning open models like OpenAI GPT or Cohere.

Step 2: Activate Readiness

- **Data Cleaning and Enrichment**
  - Use tools like Python pandas or Alteryx for data preparation.
  - Remove duplicates and outliers.
  - Fill missing values with predictive imputations where feasible.
- **Data Labeling**
  - Use manual labeling for high-value datasets or services like Labelbox.
  - Implement semi-automated labeling tools where possible.
- **Data Integration**
  - Use ETL (Extract, Transform, Load) pipelines to consolidate data.
  - Popular tools: Apache NiFi, Talend, or Azure Data Factory.
- **Real-time Data Processing**
  - Implement basic streaming capabilities if near-real-time insights are required.
  - Tools: Apache Kafka or AWS Kinesis.

Step 3: Scaling Capabilities

- **Develop a Data Maturity Roadmap**
  - Phase 1 (0-6 months): Focus on foundational data collection and organization.
  - Phase 2 (6-12 months): Deploy data pipelines and train models for simpler use cases (ChatBot, Sentiment Analysis).
  - Phase 3 (12+ months): Expand to advanced analytics and personalization engines.
- **Monitor and Improve Data Quality**
  - Establish regular audits and dashboards to track data freshness and quality.
  - Tools: Data observability platforms like Monte Carlo or Bigeye.
- **Iterate and Retrain**
  - Develop a feedback loop where model outputs inform data improvements.
  - Schedule periodic retraining of models to incorporate new data.

GLOSSARY

Term	Definition
Artificial Intelligence (AI)	A branch of computer science focused on creating systems capable of performing tasks that require human intelligence, such as learning, reasoning, and problem-solving.
Machine Learning (ML)	A subset of AI involving algorithms that enable computers to learn from and make decisions based on data without being explicitly programmed.
Natural Language Processing (NLP)	A field within AI that focuses on the interaction between computers and human language, enabling machines to understand, interpret, and generate text or speech.
Computer Vision (CV)	A domain of AI that teaches machines to interpret and analyze visual information, such as images and videos.
Robotic Process Automation (RPA)	Technology that uses software robots to automate repetitive, rule-based tasks, improving efficiency and reducing human error.

Language Learning Model (LLM)	AI systems designed to understand and generate human-like text, supporting tasks such as document summarization, chatbots, and knowledge management.
Customer Sentiment Analysis	The process of using AI and NLP to analyze customer feedback from sources like social media, surveys, and reviews to understand their emotions and satisfaction levels.
Predictive Analytics	A data-driven approach leveraging statistical models and AI to predict future outcomes and behaviors based on historical data.
Scalability	The ability of a system or process to handle increasing amounts of work or adapt to growing demands without losing efficiency or functionality.
API (Application Programming Interface)	A set of rules and protocols that allows different software applications to communicate and interact with each other.
Data Cleaning	The process of identifying and correcting errors or inconsistencies in data to ensure it is accurate and usable for analysis.
Chatbot	An AI-powered software application designed to simulate human conversation, typically used for customer service and support.
Key Performance Indicators (KPIs)	Quantifiable measures used to evaluate the success of an organization, team, or project in achieving specific objectives.
Cloud Computing	The delivery of computing services such as servers, storage, databases, and AI tools over the internet to provide faster innovation and flexible resources.
Generative AI	AI systems, such as GPT models, that can generate new content like text, images, or code based on the input provided.
Data Privacy	Measures and regulations designed to protect personal data from unauthorized access or misuse, ensuring compliance with laws such as GDPR.
Automation	The use of technology to perform tasks with minimal human intervention, increasing efficiency and accuracy.

AI ICONS

CORP STRATEGY ICONS