**CHAPTER 2**

**METHODOLOGY**

**2.1 Data Collection**

Data collection is a critical aspect of developing AI that can learn to walk autonomously, as illustrated by recent advancements in deep reinforcement learning (DRL). The process begins by gathering vast amounts of data from the robot's interactions with its environment. In traditional methods, these interactions occur in simulated environments, where virtual models of the robot and its surroundings are created to avoid physical damage and allow rapid iteration. However, simulations often fail to capture the complexities of real-world dynamics, necessitating extensive adjustments when transferring learned behaviors to actual robots.

To address these limitations, researchers have shifted towards collecting data directly from real-world interactions. This approach involves equipping robots with sensors to gather information about their surroundings and their own movements. Key data points include position, velocity, force, and environmental feedback, which are crucial for the robot to understand and adapt to its physical context. For instance, a study by Google Research focused on a four-legged robot learning to walk by collecting real-time data from its onboard sensors, which informed the reinforcement learning algorithms.

The process starts with setting up a controlled environment where the robot can safely explore and interact. Researchers use a combination of cameras, motion capture systems, and onboard sensors to monitor the robot's movements and the impact of different terrains on its locomotion. The data collected includes the robot's gait patterns, joint angles, and the forces exerted during each step. This information is continuously fed into the learning algorithm, which iteratively adjusts the robot's actions to improve stability and efficiency.

One significant challenge in real-world data collection is managing the robot's failures during the learning phase. Unlike simulations, where resets are instantaneous, physical robots require manual intervention when they fall or encounter obstacles.

Researchers mitigate this by implementing safety measures and constraints within the learning environment. For example, bounding the area within which the robot can move prevents it from wandering off and getting damaged, while additional algorithms help the robot recover from falls autonomously.

The collected data undergoes preprocessing to filter out noise and irrelevant information, ensuring that the learning algorithm focuses on meaningful patterns. This step often involves normalizing the data, synchronizing sensor inputs, and segmenting it into manageable chunks for analysis.

Researchers also augment the dataset by varying the environment, introducing different surfaces, inclines, and obstacles to teach the robot adaptability. This diversity in data helps the robot generalize its learned behaviors to new and unforeseen scenarios.

In addition to environmental data, researchers collect data on the robot's internal states, such as battery levels and motor temperatures, which influence its performance. This holistic approach ensures that the AI system can account for both external and internal factors during the learning process.

Overall, the data collection phase is integral to developing robust and adaptive AI systems capable of autonomous locomotion. By leveraging real-world interactions and comprehensive sensor data, researchers are pushing the boundaries of what autonomous robots can achieve, moving closer to creating machines that can navigate complex, dynamic environments without human intervention. This progress not only advances the field of robotics but also opens up new possibilities for deploying autonomous systems in diverse applications, from search and rescue missions to everyday tasks in human environments.

**2.2 Data Preprocessing**

Data preprocessing is a crucial step in the development of AI systems that learn to walk autonomously, ensuring that the raw data collected from sensors and interactions is refined and structured for effective learning.

Initially, the raw data often contains noise, inconsistencies, and irrelevant information that can hinder the learning process. Therefore, preprocessing involves several key steps to clean and organize the data.