Name of student : FAKIR Yasmine

Name of your level 1: IGUEME Merveille

Title: CodeGeeX: A Pre-Trained Model for Code Generation with Multilingual Benchmarking on HumanEval-X

Key words : CodeGeeX, AI model, Code generation, Code translation, HumanEval-X benchmark, Programming languages, Code explanation, IDE integration, Open source, GitHub, Developer community, Multilingual code manipulation, Code evaluation, Visual Studio Code, JetBrains, Code efficiency, Collaboration, Model parameters, Python, JavaScript

Summary:
CodeGeeX is a powerful multilingual code generation model introduced by researchers from Tsinghua University and Zhipu.AI. The model, boasting 13 billion parameters, is pre-trained on a massive code corpus comprising 23 programming languages, totaling 850 billion tokens as of June 2022. Unlike its predecessors, CodeGeeX stands out by supporting code explanation and translation, in addition to code generation.

Key Points:
Model Architecture: CodeGeeX adopts a 39-layer transformer decoder with a top query layer, following the GPT paradigm. The model is designed for autoregressive language modeling, demonstrating efficiency in code generation, translation, and explanation tasks.

Training Data: The training corpus, derived from open-source datasets like The Pile and CodeParrot, includes 23 programming languages. Each segment is tagged with a language-specific identifier to enhance the model's multilingual understanding.

Training Efficiency: The model is trained on a cluster of Ascend 910 AI processors, facing and overcoming technical challenges specific to this hardware. Training optimizations, such as kernel fusion and auto-tuning, significantly improve efficiency.

Fast Inference: To support fast inference, CodeGeeX undergoes post-training quantization, reducing memory consumption. Further acceleration is achieved through the implementation of FasterTransformer, resulting in efficient and memory-friendly deployment.

HumanEval-X Benchmark: The researchers introduce the HumanEval-X benchmark, addressing the need for a multilingual evaluation standard. It comprises 820 hand-written problem-solution pairs, spanning C++, Java, JavaScript, Go, and Python. Each pair supports both code generation and translation tasks.

Real-world Extensions: CodeGeeX is extended to popular Integrated Development Environments (IDEs) like Visual Studio Code, JetBrains, and Tencent Cloud Studio. Tens of thousands of active users benefit from diverse functions such as code completion, generation, translation, and explanation.

User Study Results: A user study reveals that CodeGeeX significantly improves coding efficiency for 83.4% of its users.

Public Accessibility: CodeGeeX has been publicly accessible since September 2022. The researchers have open-sourced the model's code, weights, API, extensions, and the HumanEval-X benchmark on GitHub.

In conclusion, CodeGeeX emerges as a robust and efficient multilingual code generation model, showcasing superiority in various tasks over existing models. Its open accessibility and compatibility with real-world IDEs position it as a valuable tool for developers worldwide.

AI Model Used: CodeGeeX, Transformer Decoder, GPT (Generative Pre-trained Transformer), Ascend 910 AI processors, HumanEval-X Benchmark, FasterTransformer, Visual Studio Code, JetBrains, Tencent Cloud Studio are some of the AI models and tools used in the project.