

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan hasil temuan dari penelitian ini dapat disimpulkan sebagai berikut :

1. *Large Language Model (LLM)* berbasis *Llama2* dapat dilakukan penyesuaian atau *fine-tuning* terhadap tugas spesifik seperti perbaikan kesalahan penulisan sintaks kode program. Tiga model berbasis *Llama2* yaitu *Llama2 Chat*, *Code Llama*, dan *Code Llama Instruct* dengan varian 7B parameter, dapat dilatih dengan baik. Saat dilakukan pelatihan, Model berbasis *Llama2* menghasilkan nilai loss yang stabil dan seimbang baik pada data pelatihan (train loss) maupun data evaluasi (evaluation loss). Selain itu, Dengan penambahan metode *Parameter Efficient Fine-Tuning (PEFT)* dan *QLoRA* pada model berbasis *Llama2*, dapat mengefisiensi waktu pelatihan menjadi lebih cepat tanpa mengurangi kinerja model.
2. Pengujian menunjukkan bahwa model berbasis *Llama2* mampu mengimbangi model sebelumnya dalam skor akurasi perbaikan program, Walaupun model sebelumnya mendapatkan hasil uji akurasi lebih baik dengan Model berbasis *Llama2*, Model berbasis *Llama2* seperti *Llama 2 Chat*, *Code Llama*, dan *Code Llama Instruct* dengan varian 7B, mendapat hasil skor akurasi dengan rata-rata 93% dalam memperbaiki kode JavaScript. *Code Llama* mendapatkan akurasi paling baik diantara model berbasis *Llama2* lainnya dengan hasil 94.2% skor akurasi perbaikan kode program.
3. Pada evaluasi kualitas kode yang dihasilkan model berbasis *Llama2* mendapatkan nilai rata-rata uji skor BLEU 72 dan rata-rata skor CodeBLEU 58 yang mengindikasikan bahwa kualitas kode yang dihasilkan oleh model tergolong sangat baik dan kode dapat dipahami oleh manusia. Model berbasis *Llama2* menunjukkan kemampuan dalam mengerjakan tugas perbaikan kesalahan secara efektif dan berpotensi

menjadi langkah maju dalam pengembangan *Large Language Model* (LLM) untuk perbaikan program otomatis.

5.2 Saran

Meskipun model yang dihasilkan telah mendapatkan nilai pengujian yang cukup baik, beberapa saran dan rekomendasi berikut dapat dijadikan sebagai bahan pertimbangan untuk meningkatkan hasil uji dan mengarahkan penelitian ini ke tahap berikutnya :

1. Meskipun telah diformat sebelumnya, Dataset perlu penyesuaian dan pembersihan tambahan dikarenakan beberapa data yang diubah ke *prompt* instruksi perbaikan kode menjadi ambigu dan sulit dipahami oleh *Large Language Model* (LLM) dengan pendekatan *Supervised Fine Tuning*.
2. Perlu mengurangi cakupan konteks kode pada perbaikan format prompting dikarenakan setelah dianalisis lebih lanjut, format prompting masih terlalu umum bagi model untuk memahami dengan tepat apa yang harus dilakukan. Untuk mempersempit konteks kode, disarankan untuk menyediakan informasi spesifik mengenai variabel atau fungsi yang bermasalah.
3. Menerapkan pendekatan metode *Graph Instruction Tuning* untuk melatih *Large Language Model* (LLM) pada sebuah kode program. Hal ini dikarenakan sebuah kode program setiap fungsi yang dibuat memiliki hirarki hubungan layaknya pohon. Oleh karena itu pendekatan *Graph Instruction Tuning* dapat dilakukan agar model dapat lebih memahami struktur kode yang diberikan.
4. Menerapkan sistem perbaikan otomatis pada *Integrated Development Environment* (IDE) dengan tujuan untuk melakukan pengujian secara langsung oleh pengembang perangkat lunak dalam mengukur efektivitas kode perbaikan yang dihasilkan oleh model berbasis *Llama2* dalam menangani *error* dan *warning*.