

## BAB 5 SIMPULAN DAN SARAN

### 5.1 Simpulan

Berdasarkan proses, implementasi program, dan evaluasi yang telah dilakukan di penelitian ini, dapat disimpulkan sebagai berikut:

1. Penelitian *benchmarking* kontrak cerdas yang dihasilkan oleh kecerdasan buatan untuk DeFi telah berhasil diselesaikan. Hasil implementasi dan *benchmarking* yang telah dilakukan menunjukkan bahwa model Code-LLaMa dan Code-LLaMa - Python dapat menghasilkan kontrak cerdas Solidity secara otomatis untuk beberapa skenario DeFi dan juga dapat memenuhi sebagian besar kriteria yang diterapkan.
2. Program *benchmarking* dapat mengevaluasi kontrak cerdas Solidity yang dihasilkan oleh kedua model tersebut. Didapatkan juga hasil berupa skor berdasarkan kriteria yang diimplementasi ke dalam program tersebut.
3. Berdasarkan proses *benchmarking* yang sudah dijalani, didapatkan skor terbesar dari masing-masing skenario DeFi dan model yang unggul untuk penghasilan kontrak cerdas Solidity secara otomatis. Skenario 1 dengan skor rata-rata 52,44 dari 20 kontrak cerdas diperoleh model Code-LLaMa, skenario 2 dengan skor rata-rata 39,77 dari 20 kontrak cerdas diperoleh model Code-LLaMa - Python, dan skenario 3 dengan skor rata-rata 39,0 dari 20 kontrak cerdas diperoleh model Code-LLaMa - Python.

### 5.2 Saran

Berdasarkan hasil dan evaluasi dari penelitian ini, terdapat beberapa faktor yang sudah didiskusikan pada 4.4. Namun, didapatkan beberapa saran utama untuk peneliti yang meneliti di bidang ini, yaitu antara lain:

1. Masalah pada program tersebut adalah hasil kontrak cerdas yang cenderung masih belum sesuai dan membutuhkan jumlah waktu cukup lama untuk menghasilkan 20 kontrak cerdas dalam satu sesi. Disarankan penggunaan ukuran jumlah parameter model Code-LLaMa yang lebih besar daripada

7B atau menggunakan model lainnya untuk memaksimalkan kualitas penghasilan *source code* kontrak cerdas Solidity untuk DeFi.

2. Skor yang dihasilkan oleh program *benchmarking* ini cenderung belum efisien, terutama pada pembobotan nilai. Seluruh kriteria ditentukan dengan masing-masing 0,1 sehingga mengubah atau meningkatkan sistem kriteria dan penilaian bobot di program *benchmarking* seperti mencoba skenario DeFi lainnya atau mencari faktor lainnya yang relevan dengan kriteria untuk memberikan evaluasi dan skor yang akurat dan optimal pada hasil generasi kontrak cerdas yang dihasilkan oleh LLM.
3. Penelitian ini hanya meneliti dan mengevaluasi kontrak cerdas yang dihasilkan oleh model, tetapi tidak menguji fungsionalitasnya pada testnet dikarenakan yang diteliti hanya apakah model AI dapat menghasilkan kontrak cerdas Solidity atau tidak. Disarankan untuk menguji fungsionalitasnya pada penelitian selanjutnya untuk *benchmarking* kontrak cerdas yang lebih efektif.

