

A Benchmark for Recipe Understanding in Autonomous Agents

Robin De Haes

"Can a machine understand an everyday activity, like cooking, deeply enough for successful execution?"

MUHAI Recipe Execution Benchmark

- Representation language
- Kitchen simulator
- Suite of multiperspective metrics
- Recipe dataset

Open Source: <https://ehai.ai.vub.ac.be/recipe-execution-benchmark>

Gold Standard Dish	Predicted Dish
baking-tray <i>located at counter-top</i> <ul style="list-style-type: none">lined-with: baking-paperarrangement: side-to-side <i>25 portions</i> homogeneous-mixture <ul style="list-style-type: none">temperature: 175 °Ccurrent-shape: crescentbaked: truemixing-type: mixedamount: 25 g all-purpose-flour <ul style="list-style-type: none">sifted: falseamount: 10 g white-sugar <ul style="list-style-type: none">temperature: 18 °Camount: 5 g butter <ul style="list-style-type: none">temperature: 18 °Camount: 10 g	cookie-sheet <i>located at counter-top</i> <ul style="list-style-type: none">lined-with: baking-paperarrangement: side-to-side <i>20 portions</i> homogeneous-mixture <ul style="list-style-type: none">temperature: 175 °Ccurrent-shape: ballbaked: truemixing-type: beatenamount: 36.25 g cocoa-powder <ul style="list-style-type: none">amount: 5 g all-purpose-flour <ul style="list-style-type: none">sifted: falseamount: 13.125 g white-sugar <ul style="list-style-type: none">temperature: 18 °Camount: 6.25 g butter <ul style="list-style-type: none">temperature: 5 °Camount: 13.125 g