The Many Challenges of Human-Like Agents in Virtual Game Environments

Maciej Świechowski, May 23, 2025, Detroit

&

Dominik Ślęzak

About me

Industry, Foundation, and Academia (PhD in 2015)

At this conference: representing a company

- software services for intelligent agent development
- tools for the gamedev industry

This paper = two parts

Part 1: Mini survey about the challenges of creating human-like agents.

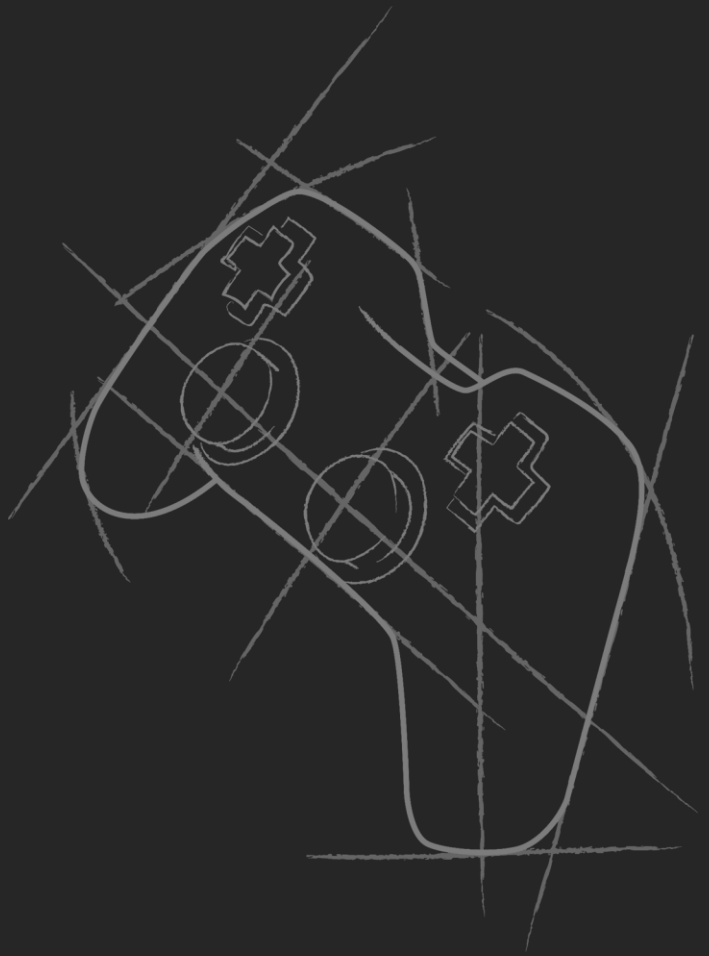
- 13 challenges
- 54 filtered research papers

Part 2: Empirical experiment in our own game – creating ML model to distinguish humans from bots.

Infeasible to cover all topics → this presentation will only be a teaser.

Part 1: The challenges

(included in the paper)



(1) Humans are Diverse

- *“behaving in a manner that makes it indistinguishable from human players”*
- *“giving the feeling of being controlled by a (human) player”*

- **ambiguity and imprecision**

- **many categories of players** (listed in the paper)
- skill level is the most crucial factor distinguishing different playing patterns

(2) The Complexity and Expressiveness of Action Space

Complex high-dimensional action spaces:
a great challenge in implementing human-like AI

Simple, constrained environments: complicate the evaluation of whether an AI is truly human-like. Example: *tic-tac-toe*.

Environment must possess a sufficient level of expressiveness to enable feasible assessments of human-like behavior.

(3) The Challenges of Scale

Even one human-like character can be computationally demanding

Now, imagine entire city!

“Crowd AI” - uses simplified techniques

(4) Avoiding Superhuman Behavior

Rationally acting towards goals in a smart way is believable.

However, too high efficacy in achieving goals => superhuman

Superhuman is not believable.

Examples:

- superhuman reflexes
- superhuman calculations
- superhuman precision (e.g. of shooting)

(5) Idle and Non-Relevant Actions

Humans often spontaneously engage in behaviors irrelevant from the perspective of the goal.

"*for fun*" or "*for own amusement*"

(6) Biological Constraints

It is hard to simulate biological constraints in a realistic fashion.

Examples:

- **visual perception**
- **sound perception**
- **memory**
- **cognitive load**
- **ability to multitask**
- **reaction time**
- **fluctuation, delay, tiredness**

(7) Emotional Element

In literature, displaying emotions listed as one of the most critical qualities of believable bots after “goals”.

Examples:

- **emotional response to events**
- **different behavioral patterns based on emotions**
- **compassion**
- **anger**
- **tilt**

(8) Handling Uncertainty

Typically involves hidden information, randomness or both.

Introduces additional layer of complexity.
- in video games, bots are often given unfair access to information.

Under uncertainty, video game agents are typically either:
- **Incompetent**
- **Strong but cheating**

Solution is expensive:
- human-like reasoning
- inferring missing information
- probabilistic thinking

(9) Adaptability

AI in commercial games is often exploitable due to its tendency to follow repetitive patterns ([28] in the paper).

Adaptation to other agents (e.g., opponents) and scenarios.

(10) Making Mistakes

Humans sometimes make mistakes.
Humans often learn from their mistakes.

Implementing both is challenging in complex real-time environments.

“An intelligent agent model should not require producing a “perfect” agent, but rather, for better human resemblance and higher believability, it is more natural to have the flaws and dysfunctionalities of the human affect phenomena incorporated into the model.” ([12] in the paper)

(11) Training Human-Like AI (using Machine Learning)

We want human-likeness => training from human data.

We share some existing challenges:
(1) human are diverse, (2) computational demand, (3) the effect of scale

And additionally:
- large volumes of training data are required
- lack of data during game development or if a game is unpopular
- human-like control input:
    - using controllers versus “move to point (X,Y)”

(12) Simulating Social Norms

- frequently mentioned in the literature

In the paper, we refer to 3 bibliographic sources and the game:
The Elder Scrolls IV: Oblivion by *Bethesda Software*.

(13) Specific Human-Like Activities in the Game

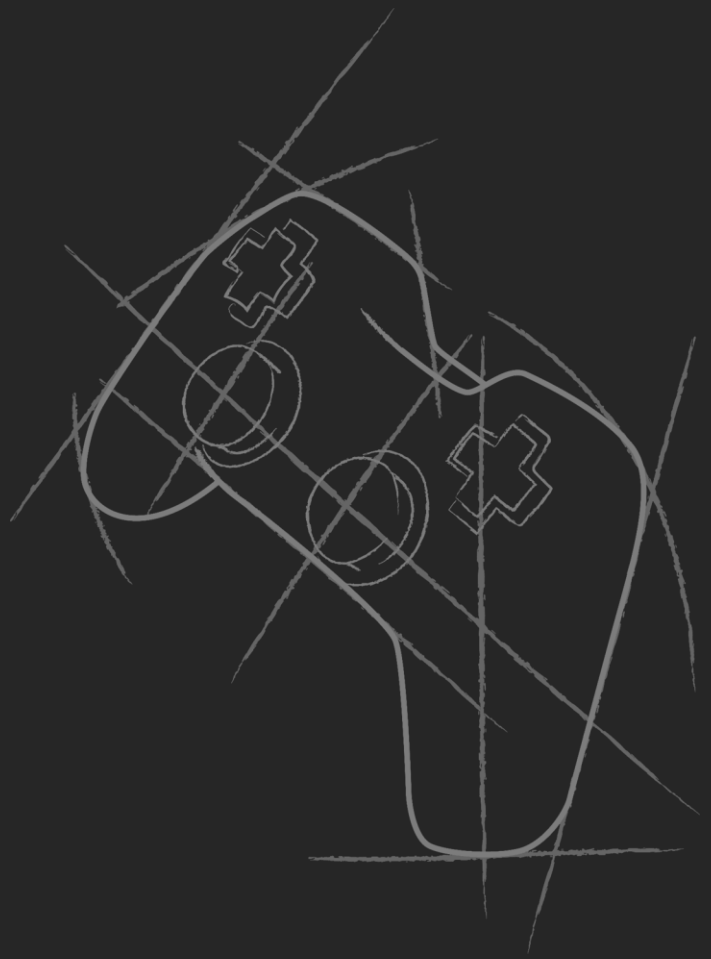
Games (and other virtual environments) may include specific activities such as:

- driving a car
- walking in a city
- conversing in a natural language
- building structures

Each such activity can be assessed for their human-likeness.

Example: *Turing Test* for conversation.

Part 2: Experiment



The Game and Data

Tactical Troops: squad-based top-down tactical game (“XCOM meets Worms”)

Detailed game logs.

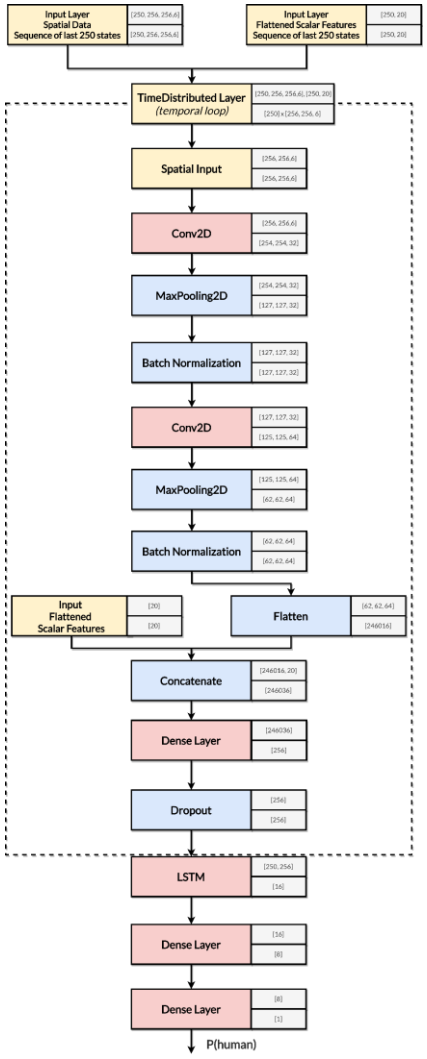
- **bot vs. bot:** 89667
- **human vs human:** 1338
- **human vs bot:** 2190

The Model

RNN + CNN
(20 features + 6 maps (2D)) x 250 last states

Model	F1-Score	Accuracy
XGB	0.58	0.68

Model	F1-Score	Precision (Human)	Recall (Human)	Cohen Kappa
Full →	**0.92**	**0.87**	**0.81**	**0.84**
RNN only	0.88	0.75	0.79	0.78
CNN only	0.59	0.61	0.57	0.58



Motivation for creation & detection methods

In game industry:

- **Model as QA during game development**
- Bot detection
- Overcoming the “cold start” in MMO games
- Elevating entertainment
- More accurate testing (testing what humans might do)

Beyond game industry:

- Fake information detection
- Identify verification
- Measuring the progress of AI
- Next-generation CAPTCHA

Conclusions

The creation of human-like bots and the measurement of their human-likeness each have distinct advantages.

Developing human-like agents presents a significant challenge.

Differentiating bots from humans can be achieved with relatively high accuracy in quite complex game environments.

The complexity of creating human-like agents in a specific environment is inversely correlated with the ease of developing methods to distinguish between humans and bots in that environment.

Thank you!

maciej.swiechowski@qed.games

https://grail.com.pl/

Extra slide about the model

Numerical features [x2, for both players]:
turn number, damage dealt, damage received, friendly-fire damage, friendly-fire to total damage ratio, # used grenades, damage dealt using grenades to total damage, # used gadgets, # of units' status changes

Spatial features:
friendly unit positions with health (health values represented as color saturation), enemy unit positions with health, obstacles, rooftops, teleports, control points.

This research was co-funded by the Smart Growth Operational Programme 2014-2020, financed by the European Regional Development Fund under a GameINN project POIR.01.02.00-00-0207/20, operated by The National Centre for Research and Development