Nudging Robot Engineers To Do Good: Developing Standards for Ethical AI and Robot Nudges

Ken Bell[1], Sean Dougherty[2], Vivek Nallur[3], John P. Sullins[4]

1- IEEE P7008 Working Group and BDO Canada LLP; 2-IEEE P7008 Working Group; 3-University College Dublin; 4-Sonoma State University

P7008
A/IS Nudging Standards IEEE

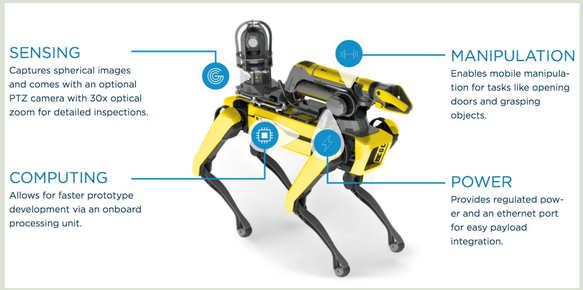
Introduction

A new kind of AI system is rapidly becoming more common. HyperNudging or A/IS (Autonomous Intelligent Systems) Nudging allows programmers to engage in changing the behavior of users, as opposed to simply predicting the possibility of a user engaging in a desired action. A/IS Nudging systems are new but they are already recommending products and entertainment options that are autonomously curated by the system for specific users. We wear systems on our body that encourage us to walk an extra thousand steps, or remind us to get up and stretch even though we have a conference abstract due. Algorithms attempt to keep us engaged with a particular social media platform for as long as possible, by queueing up endless streams of content chosen specifically for us. In the near future we will begin to see more robot systems doing similar things in the physical world. Examples include robots to encourage safe behavior in parks, stores or malls, to help officials monitor public health or enforce quarantines, and to invite people to linger in shops, museums and malls. There are already healthcare robots being developed for the elder-care domain – where robots remind elderly residents to take their medicine, or talk to their family, or go for a walk

In this poster, we describe two use case scenarios of A/IS Nudging and show how the IEEE P7008 standard is being designed to help engineers build systems more attuned to ethical outcomes than would be found in the absence of any standards guidance. Example one describes the use of A/IS nudging to detect the use of gendered nouns and to try different nudges to help the human use non-stereotyped language. The second example looks at A/IS robotic systems used to attempt to enforce social distancing and mask use during the COVID pandemic. Our poster will include our thinking on the types of standards required for applications, such as our examples, and on the specific ethical and legal issues for which we are tasked with providing guidance.

Objectives

1. Discuss the current status of the IEEE P7008 - Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems standards being developed for encouraging ethical outcomes for AI and Robot Nudges.
2. Describe two use case scenarios of A/IS Nudging and show how the IEEE P7008 standards are being designed to help engineers build systems that are more attuned to producing more ethical outcomes then would be found in the absence of any standards guidance.
3. Encourage our colleagues in AI and Robotics Law to consider joining our committee which is nearing completion of writing our standard



Example 1

A/IS Nudging to reduce the use of Gender Stereotypes in Language

Many languages have gendered aspects built into the language itself. For instance, in German nouns change their form depending on the gender of the person. This can be an issue when designing A/IS applications such as healthcare robots or customer service chatbots, as they tend to perpetuate gender stereotypes that many societies would like to move beyond.

Aim of the project: Detect the use of gendered nouns and try different nudges to help the human use non-stereotyped language.

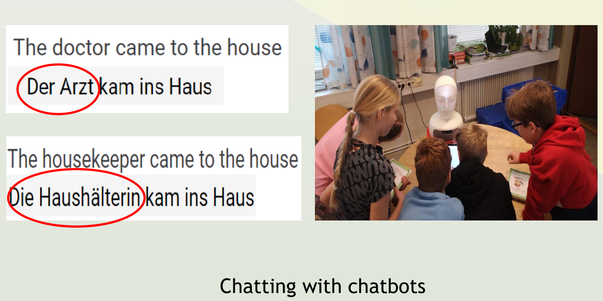
Device contexts: a) Smartphone; b) A conversational robot that interacts with students

Desired outcome(s): Awareness of gender stereotyped nouns [e.g. doctor/engineer - male, housekeeper - female] and selection of gender-neutral language where appropriate

Ethical challenges: Obstruction of freedom of expression

Technological challenge: Detecting when to nudge [sometimes a gender-neutral noun is incorrect, in context], and detecting whether the nudges were having an effect

Nudge Design: In smartphones, offer auto-completion with gender-neutral nouns as the default option, with the ability for the user to choose other nouns. In the conversational robot, probabilistically tilt towards discussing career options that are not typically chosen by the gender of the student.



Chatting with chatbots

Spot in Singapore

Example 2

A/IS robotic systems used to attempt to enforce social distancing and mask use during the COVID pandemic.

A/IS Nudging for encouraging social distancing and mask use during the COVID Pandemic.

Aim of system: Used by police and security to detect members of the public who may be ignoring recommendations to socially distance or wear masks in public areas such as parks.

Device Contexts: A) 4 legged robotic systems developed by Boston Dynamics. Spot Explorer approx. $75,000 per machine. B) speakers and cameras mounted on the system so the officers who are teleoperating the system can interact with the public in safety. “The robot isn’t really enforcing in Singapore,” Boston Dynamics founder Marc Raibert told CNBC’s “Squawk Box” last May. “It’s just giving people information and encouraging them,” he said. “There’s a human nearby who can do whatever enforcement they decide is appropriate” (Matthews, 2020). C) system used in Singapore and in NYC

Desired outcome(s): Increased compliance with mask and social distancing mandates to help lower infection rates.

Ethical challenges: Not well received causing a negative PR problem in NYC but largely tolerated in Singapore. Highlights the question on whether the technology is unethical, or violates civil liberties (Bushwick, 2021).

Technological challenge: Can small robots act in effective ways to nudge compliance with public health mandates and suggestions?

Nudge Design: A Boston Dynamics four legged robot was used due to its availability, its design facilitated moving across terrain found in public parks

Spot in Germany (The Telegraph)

Spot in Honolulu

P7008 Principles that could impact these Two Examples

1. **Beneficence:** Increased well-being of human beings as primary success criterion of nudge development [Principle #2 of EAD, e1]
2. **Transparency:** The identity of the parties nudging the user can be disclosed. The user is aware that they will be nudged by the system.
3. **Human Rights:** A/IS nudges will not be built in ways that infringe on human rights. Nudge designers will show due diligence in respecting these user rights (as described in UN, The Human Rights Council, 2011).
4. **Explicit Contract:** Opt-in system policy with explicit consent**.** 1 - Contract for designer (formal code of ethics /contract with reference to P7008). If the AI goal is persuasion, a design responsibility is created. **2 -** End-user license for explicit consent

References

Bushwick, Sophie, The NYPD’s Robot Dog Was a Really Bad Idea: Here’s What Went Wrong: Design, context and timing influence whether humans embrace a robot or reject it, Scientific American, May 7, 2021

Matthews, Korey, Boston Dynamics’ dog-like robot can now recharge on its own and gains popularity in Covid era, October 21, 2020

WBUR News, Boston Dynamics Robotic Dogs Used By Honolulu Police To Scan Homeless People For Fevers, July 30, 2021

Morris, Chris, Bad dog! NYPD kennels its robotic K9 after backlash, Fortune, April 29, 2021.

Schmidt, Andreas T. and Engelen, Bart, The ethics of nudging: An overview, Philosophy Compass. Wiley, ;15:e12658, 2020

UN, The Human Rights Council, Guiding Principles on Business and Human Rights, 2011

Zaveri, Mihir, N.Y.P.D. Robot Dog’s Run Is Cut Short After Fierce Backlash, New York Times, April 28, 2021

How to Get Involved!

IEEE P7008™

Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems

IEEE Standards Project for Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems establishes a delineation of typical nudges (currently in use or that could be created) that contains concepts, functions and benefits necessary to establish and ensure ethically driven methodologies for the design of the robotic, intelligent and autonomous systems that incorporate them. “Nudges” as exhibited by robotic, intelligent or autonomous systems are defined as overt or hidden suggestions or manipulations designed to influence the behavior or emotions of a user. **Join**