

Robots as AI Double Agents: Privacy in Motion Planning

Rahul Shome, Zachary Kingston, and Lydia E. Kavraki

Abstract—Robotics and automation are poised to change the landscape of home and work in the near future. Robots are adept at deliberately moving, sensing, and interacting with their environments. The pervasive use of robotics promises societal and economic payoffs due to its capabilities—conversely, the capabilities of robots to move within and sense the world around them is susceptible to abuse. Robots, unlike typical sensors, are inherently autonomous, active, and deliberate. Such automated agents can become *AI double agents* liable to violate the privacy of coworkers, privileged spaces, and other stakeholders. In this work we highlight the understudied and inevitable threats to privacy that can be posed by the autonomous, deliberate motions and sensing of robots. We frame the problem within broader sociotechnological questions alongside a comprehensive review. The privacy-aware motion planning problem is formulated in terms of cost functions that can be modified to induce privacy-aware behavior: preserving, agnostic, or violating. Simulated case studies in manipulation and navigation, with altered cost functions, are used to demonstrate how privacy-violating threats can be easily injected, sometimes with only small changes in performance (solution path lengths). Such functionality is already widely available. This preliminary work is meant to lay the foundations for near-future, holistic, interdisciplinary investigations that can address questions surrounding privacy in intelligent robotic behaviors determined by planning algorithms.

I. INTRODUCTION

Recent advances have introduced robots—particularly robots with manipulation capabilities—into applications such as home assistance [1], healthcare [2], service [3], and industry [4]. These settings require the robot to (i) *adapt* to sensed information (e.g., camera images) which is probabilistic and uncertain and (ii) share space and interact with humans, introducing ethical concerns. We contend that the powerful capabilities of these systems urgently burden us with novel ethical concerns relating to unprecedented use of these systems which, if not addressed now, will lead to dystopian uses of robotics by naïve or malicious actors. Robots are inherently tools of surveillance, have unprecedented access to spaces [5], are trusted in ways cameras and other technology is not [6], and have sensing capabilities that are poorly understood [7]—a robot that avoids you is also tracking you. Modern uses of robots combine consumer hardware with intricate frameworks of open-source libraries, middleware [8], learned models, and probabilistic algorithms—all of which exacerbate the opacity of a robotic system. Potential abuses are understudied, under-litigated [9] and traditional mitigation

The authors are affiliated to the Department of Computer Science, Rice University, USA. RS is now affiliated to the School of Computing at the Australian National University, Canberra. rahul.shome@anu.edu.au, {zak, kavraki}@rice.edu. This work was supported in part by NSF 1718478, NSF 2008720, and Rice University Funds.

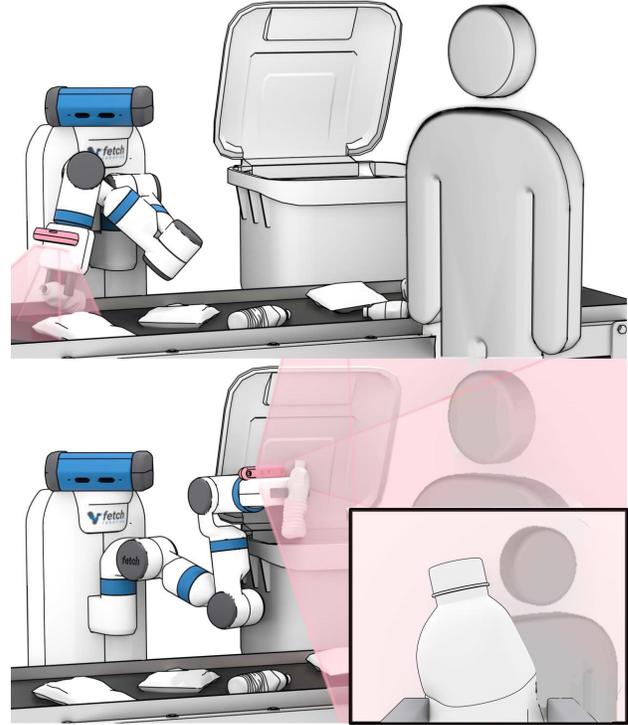


Fig. 1. Deployments of robots with sensors can expose threats to privacy from the ability of robots to autonomously use motion planning for choosing how its sensors gather data. In the figure, a sensor attached to a manipulator can gather data about coworkers during typical operation.

strategies are hard or impossible to apply, motivating an urgent need for understanding such threats.

Privacy violations by robots can take many forms, such as data over-collection beyond what is strictly necessary for operation that can be used for later inference (as is often observed in web technologies [10], [11]). Compromised robotic systems with poor security can leak information to unknown third parties [12], [13]. Exfiltration of sensor data or post-hoc analysis of camera data can violate privacy by monitoring coworkers or users, extracting privileged information from the workplace and the home (e.g., [14]). It behooves us to be wary of the uneasy parallels between the proliferation of robots for generating economic value and surveillance capitalism [15].

Focus in the literature has primarily addressed privacy concerns raised by robots and smart devices from a computer vision perspective [16]–[19]. There is little work from a privacy standpoint on what makes robots unique—their *ability to move and interact with the physical world*. In this work we focus on a fundamental element of robotic autonomy—

motion planning—and its relation to privacy. Robot operators can use custom costs, constraints, and objectives that may place humans co-working with robots in situations that may violate their privacy. Abuses to privacy gives rise to what we call *robots as AI double agents*. To the best of the authors’ knowledge, a comprehensive study of privacy implications of generalized motion planning is sorely under-explored.

The key contributions of this work are to present (a) a detailed motivating review of interdisciplinary literature that explains considerations of privacy at the confluence of society, technology, and engineering drawing out the connection to the robotic motion planning problem; (b) a formalization of the privacy in the motion planning problem that identifies privacy-violating functional definitions of feasibility and costs as potential vulnerabilities; (c) a set of motivating case studies based on typical manipulation and navigation scenarios using simple simulations and a straightforward weighted cost function to demonstrate that (i) simple cost function alterations can cause severe privacy-violating behavior, and (ii) privacy-violating behavior can be accompanied by only minor changes in traditional performance metrics (path length). The technical choices in the simulated study are simple modifications to the motion planning problem, using readily available open-source functionality, that serve to provide an illustrative testbed. The takeaways from this work points out the clear and present dangers to privacy posed in robotic motion planning.

II. THE BIGGER PICTURE OF PRIVACY AND ROBOTICS

We first take a step back and look at where robotics lies within the broader context of engineering and cyber-physical systems. Many of the privacy considerations attributed to traditional uses and abuses of technology are aggravated by the power of robotic systems to not only be passive sensors, but also be autonomous in the physical space.

Privacy: We must concretely define what we mean by *privacy* [20]. A precise definition is closely tied to societal and legal interpretations in different parts of the world. We choose to refer to GDPR, a push towards common law privacy safeguards [21]. A closely related definition [22] promises safeguards that “*protects the individual against abuses which may accompany the collection and processing of personal data and which seeks to regulate at the same time the transfrontier flow of personal data.*”

Note that we are separating security concerns from those of privacy (e.g., see [13], [23]–[26] for ROS and security). However, privacy violating systems are more readily exploited when security has been compromised.

Engineering Ethics: In the context of robotics, a significant portion of the system design falls on automation deployers, consultants, and engineers. This draws a close connection between the questions of ethical automation and engineering ethics [27]. Ethics has been studied as an important aspect of engineering problems where solutions have to trade off ethical considerations and risks versus profit, efficiency, and output. The choices made by engineers can have critical societal impacts and unethical choices can have significant fallout. Engineering ethics is also deeply connected to

morality and responsibility [28]. Intelligent automation lies under the shadow of this complicated relationship between technology, ethics, and society. Beyond sharing many common problems [29], the powerful capabilities of robotics presents unique challenges and threats.

Cyber-physical Systems: While analysis on privacy has been done in traditional cyber-physical systems [30], [31], the internet of things [32], and so on, intelligent robot systems have received little analysis. There also has been little understanding and observation from policy makers to the threat that robotics bring to privacy, particularly as these systems become more ubiquitous [9], [33]. There is an uncomfortable relationship between smart home devices and considerations of privacy and legalese [34], [35], including innocuous products like smart toys [36].

Robots as Sensors: From the perspective of a robot as a passive sensor platform, there has been much work in preserving privacy [37], through methods such as obfuscating sensor input [16], using degraded images (e.g., anonymizing faces [17], reducing quality of the camera feed to a teleoperator [18]), and redacting relevant parts of the scene [19]. There has also been use of “privacy markers” that indicate regions that should be removed or redacted from sensors [38], [39], automatically detecting these regions [40], and also extended to a case with mobile robots [41]. However, all these methods merely operate on the camera feed passively, and do not actively direct what information the sensor should gather. Even with minimal data, powerful inference can identify individuals (e.g., with motion [42]). In general, robots must collect only the data they need [33].

Robots around Humans: Trust is essential for embodied systems to operate reliably near humans [43]. Moreover, people are more “comfortable” with robots rather than unembodied cameras, and more willingly expose themselves to privacy violations [6], and misunderstand the full capabilities of robotic systems to gather information [7], [44]. There are certain qualities that humans expect from robots, and how that relates to how robots can “fly under the radar” when doing things. This relates to intention-aware planning [43].

Privacy and Learning: There has been much recent work on using machine learning-based methods for robot control, particularly in learning from human demonstrations [45], [46]. Learning based methods require large amounts of data, which is at odds with privacy concerns that require minimal data collection. There has been work in addressing privacy in deep learning [47], [48], namely in differential privacy [49], [50], but little from this literature has been applied to robotics and control [51], particularly in the context of manipulation.

Given the complexity of machine learning-based models, there is also potential for subterfuge, e.g., adding undetectable backdoors to modify behavior of a system [52]. Such backdoors could be used to induce malicious behavior in models used to nominally preserve privacy without the awareness of stakeholders.

Privacy in Robotics Research: Privacy is pressing issue throughout the broader AI community [53]. Robotics in

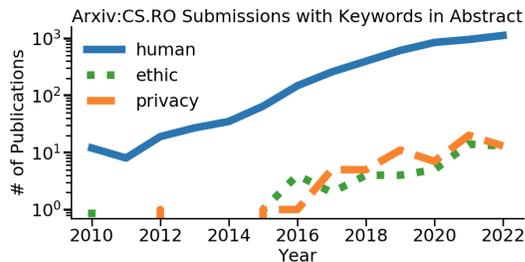


Fig. 2. The figure shows the logarithmic growth of submissions to Arxiv:CS.RO with the keywords *human*, *ethic*, and *privacy* in their abstract between 2010 and 2022. Privacy and ethical concerns are far outpaced by applications that interact with humans.

particular is a concern for privacy, given the direct nature of a robotic system as a tool of surveillance, with sensing an inherent part of a robot’s ability to understand the world [54]. However, considerations of privacy in the field of robotics have been far outpaced by the potential usages that might pose threat. Fig. 2 is an approximation of what is an undeniable trend—robotics solutions and applications that have the potential to interface with humans are growing beyond the understanding of privacy (and ethics) in these contexts. [55] makes a similar observation.

There is some work in understanding privacy concerns for social robots [56], [57]—that is, systems primarily designed for human interaction and entertainment. Systems, such as household robots, have been shown to have poor security and privacy properties [12]. We focus instead on general scenarios where the intended purpose, such as a logistics task, can be compromised by motion planning.

Privacy and Planning: There has been prior work in incorporating deliberate reasoning about sensing within planning. Generally falling into the category of “active vision” [58], work exists in autonomous surveillance [59], searching for objects [60], [61], and chronicling [62], [63]. Some work has used the capabilities of robotic arms [64] in assistive applications [65]. There has been work for protecting the privacy of robots themselves [66]–[68], but these works typically apply to mobile or aerial applications. Drones form a common platform of choice for visibility-aware problems [69] with some work on aspects of privacy [70]–[75]. These works focus on clearly defined areas to avoid (e.g., similar to explicitly marked privacy zones) but only deal with low-dimensional systems (e.g., mobile robots and drones). General robots with many joints—like manipulators that are beginning to be more broadly deployed—pose significant challenges due to the high-dimensionality of their search space. Addressing the problem for general platforms and manipulators is necessary to apply to broader scenarios.

III. PRIVACY-AWARE MOTION PLANNING

In this section we take a closer look at the role privacy plays in a fundamental aspect of robotics—motion planning. We introduce some of the technical and theoretical tools necessary to define and understand elements of privacy in motion planning. We focus on a threat model where a robot fitted with a sensor collects data while moving. Privacy-aware

motion planning will be defined in terms of data collected on privacy-sensitive regions. The vulnerabilities framed in this section can potentially lie exposed to deployers and end-users. Modifications to parameters, like cost functions, in motion planning can lead to altered behavior by robots acting as AI double agents.

Definition 1 (AI Double Agent). *An AI double agent is a robot, that by virtue of altered reasoning and planning, exhibits autonomous behavior which violates the privacy of any human agent in the robot’s workspace.*

A. Privacy-Aware Motion Planning

A robotic agent \mathbf{A} with d degrees of freedom, e.g., robotic arm with d joints, is situated in a workspace $\mathcal{W} \in \mathbb{R}^3$ with obstacle regions $o \subseteq \mathcal{W}$. The robot has a d -dimensional configuration space $\mathcal{X} \subseteq \mathbb{R}^d$. Each configuration of the robot can be checked for feasibility, typically defined in terms of being collision-free with the obstacles. Denote a boolean feasibility function as $\mathbf{v} : \mathcal{X} \rightarrow \{\mathbf{1}, \mathbf{0}\}$. The invalid subset corresponds to $\mathcal{X}_{\text{obs}} = \{x \mid \mathbf{v}(x) = \mathbf{0}, x \in \mathcal{X}\}$, while the valid subset is $\mathcal{X}_{\text{free}} = \mathcal{X} \setminus \mathcal{X}_{\text{obs}}$. A motion planning problem requires connecting a start configuration $x_0 \in \mathcal{X}$ to a goal region $\mathcal{X}_{\text{goal}}$ with a continuous, collision-free, time-parameterized trajectory $\pi : [0, 1] \rightarrow \mathcal{X}_{\text{free}}$ where $\pi[0] = x_0, \pi[1] \in \mathcal{X}_{\text{goal}}$. Given all possible such feasible trajectories $\Pi \ni \pi$, a cost function $\mathbf{c} : \Pi \rightarrow \mathbb{R}_{\geq 0}$ assigns a non-negative real number to a trajectory. The cost is typically considered to be (or some function proportional to) the Euclidean path length. An optimal solution corresponds to the minimum cost $\pi^* \in \arg \min_{\pi \in \Pi} \mathbf{c}(\pi)$.

In this work we introduce the element of privacy into the motion planning problem. For notational clarity we will use subscripts with \mathbf{p} to denote privacy-aware variants. A privacy-sensitive regions is a region of the workspace which is associated with requirements for privacy preservation is denoted by $o_{\mathbf{p}} \subseteq \mathcal{W}$. A set of k such regions is denoted by $\mathcal{O}_{\mathbf{p}} = \{o_{\mathbf{p}}^1 \cdots o_{\mathbf{p}}^k\}$. A privacy feasibility function applies to a configuration and is denoted by $\mathbf{v}_{\mathbf{p}} : \mathcal{X} \rightarrow \{\mathbf{1}, \mathbf{0}\}$. Without loss of generality we will consider privacy-violating evaluations when $\mathbf{v}_{\mathbf{p}}$ evaluates to $\mathbf{0}$. A non-negative privacy-aware cost is defined along a trajectory $\mathbf{c}_{\mathbf{p}} : \Pi \rightarrow \mathbb{R}_{\geq 0}$.

The evaluation of both the constraint and cost $\mathbf{v}_{\mathbf{p}}$ and $\mathbf{c}_{\mathbf{p}}$ will depend upon the privacy regions $\mathcal{O}_{\mathbf{p}}$. The exact nature of this relationship will be affected by the precise setting under consideration including the kinematics of the robot, the attachment of the sensor, the sensing model (for instance visibility cone for a camera, etc). Our general formulation will leave these as necessary privacy-aware pieces within the otherwise typical motion planning problem. Note that the definition of the problem thus far can be applied to many general combinations of robots, sensors, and privacy regions.

B. Types of Privacy-Awareness

The definitions of $\mathbf{v}_{\mathbf{p}}$ and $\mathbf{c}_{\mathbf{p}}$ can allow interactions with privacy regions $\mathcal{O}_{\mathbf{p}}$ with three types of privacy-awareness: **Privacy-Agnostic** (\mathbf{v}, \mathbf{c}) classical motion planning has unmodified feasibility and cost functions.

Privacy-Preserving (v_p^+, c_p^+) choices penalize privacy violations along robot motions.

Privacy-Violating (v_p^-, c_p^-) choices promote privacy violations along robot motions.

Privacy-aware behaviors present a choice of functional alternatives. This leads us to the step where these are defined, which is up to the problem designer or deployer.

Definition 2 (Motion Planning Double Agent). *The design of privacy-violating v_p^- or c_p^- to replace the privacy-agnostic feasibility and cost functions, creates a double agent generating motion plans for the privacy-violating variant of the problem.*

C. Privacy as a Secondary Objective

The privacy-aware cost function c_p , which is either privacy-preserving or violating behavior, can encode or be a part of multiple objectives [76] within the motion planning problem. The threat model of interest, and indeed of greater risk and harder to detect, is expected to involve robots that perform their primary automation operation satisfactorily while *also* achieving a secondary privacy-aware objective. For instance, consider a possible combination of the length of solution path (a traditional cost in motion planning) and the privacy region visibility by a camera attached to the robot. The manner of this combination can lead to different flavors of pareto-optimal [77] problems, while the continuous nature of the problem can relate to cost maps [78].

*Key to the broader scope is not the prescription of such a specific cost and feasibility. Rather, in the next section, we show that it suffices to design **minor changes to the cost function in standard motion planners to make them privacy-aware**. This is particularly interesting because of the relatively low barrier of access, as it might be possible for deployers or end-users with enough expertise to modify the parameters and modules of motion planning.*

IV. MOTIVATING SIMULATED CASE STUDY

In this section we demonstrate how—using a candidate modified cost function—privacy-aware behavior can be injected into normal operation of a motion planner.

A. Candidate Model of Privacy-Aware Cost Function

Consider a PRM* [79] as the motion planner that reports shortest paths over roadmaps [80] constructed in the robot’s configuration space. *Custom cost functions* change the graph edge weights, altering the discovered solutions executed by the robot. This basic functionality is readily available through powerful open-source libraries [81]. While being careful not to prescribe what a privacy-aware cost function should be, we provide a straightforward candidate for studying the effects introduced by weighted modifications to the Euclidean path length cost function. The trajectory π_p is weighted multiplicatively or fractionally using a *privacy weight* (w) parameter (such that $|w| \geq 1$) depending upon the interaction with the privacy regions. A negative weight is privacy-violating. The total cost will be calculated over a discretization ($\Delta\pi$) of the trajectory π_p . $|w| = 1$ is privacy-agnostic.

Privacy-Preserving ($w > 1$):

$$c_p^+(\pi_p) = \sum_{\Delta\pi \in \pi_p} \begin{cases} w \|\Delta\pi\| & \text{if privacy violated} \\ \frac{1}{w} \|\Delta\pi\| & \text{otherwise} \end{cases} \quad (1)$$

Privacy-Violating ($w < -1$):

$$c_p^-(\pi_p) = \sum_{\Delta\pi \in \pi_p} \begin{cases} \frac{1}{|w|} \|\Delta\pi\| & \text{if privacy violated} \\ |w| \|\Delta\pi\| & \text{otherwise} \end{cases} \quad (2)$$

Here $|\cdot|$ and $\|\cdot\|$ denote the absolute value and Euclidean arc length. In essence, all that is needed is an approximation of privacy violation (for instance, intersection with a camera cone with \mathcal{O}_p), and functional choices that penalize or promote the privacy preservation or violation. Due to the generality of the underlying planning, this should apply to a large variety of sensor-attached high-dimensional robotic platforms. Though the cost function weighting represents a simple alteration, what is not obvious is *how the cost function affects the double agent’s privacy and efficiency?*

B. Case Studies

To highlight motivating scenarios, we focus on two case studies with cameras attached to the robot while it moves. The camera visibility is approximated by a cone (shown in pink in Fig. 3 top) defined similar to the specifications of an *Intel Realsense* camera (a 42° field of view and 2m range). While sensing these privacy regions poses its own challenges, here we choose to focus on the effects of planning by assuming these as input. Our motivating scenarios will introduce typical workspace settings where specific areas might be expected to contain these privacy-regions. Humans are represented as static mesh obstacles with spherical approximations (40cm radius) of privacy regions around their heads.

Manipulation: A manipulator with a camera attached to its wrist is set up in a workspace opposite to human collaborator(s) or customer(s). These types of settings are common to *warehouse automation settings or service industries*. The task itself is typically concentrated in the shared workspace between the robot and the human, for e.g., a table, a cashier’s desk, a counter, etc. The robot is free to move the sensor (wrist) unencumbered, as long as it reaches its planning goal and avoids obstacles. The camera interacts with the regions of the workspace expected to contain human occupancy (spherical approximations shown in Fig. 3 (left, middle) for one and three individuals respectively).

Navigation: A mobile robot with a camera attached to a controllable joint is set up to navigate through a planar workflow filled with human co-workers or crowds. Such scenarios will come up in a variety of *mapping, navigation, cleaning, and monitoring* tasks where the mobile robot is operating within the floorplan while avoiding the obstacles (here humans). Since the head camera is freely controllable during its motions, the camera interacts with the regions of the workspaces expected to contain human occupancy (spherical approximations shown in Fig. 3 (right) for nine individuals).

Experimental Details: The simulation uses a Fetch in two controllable modes: a) **arm+torso**: a camera attached to

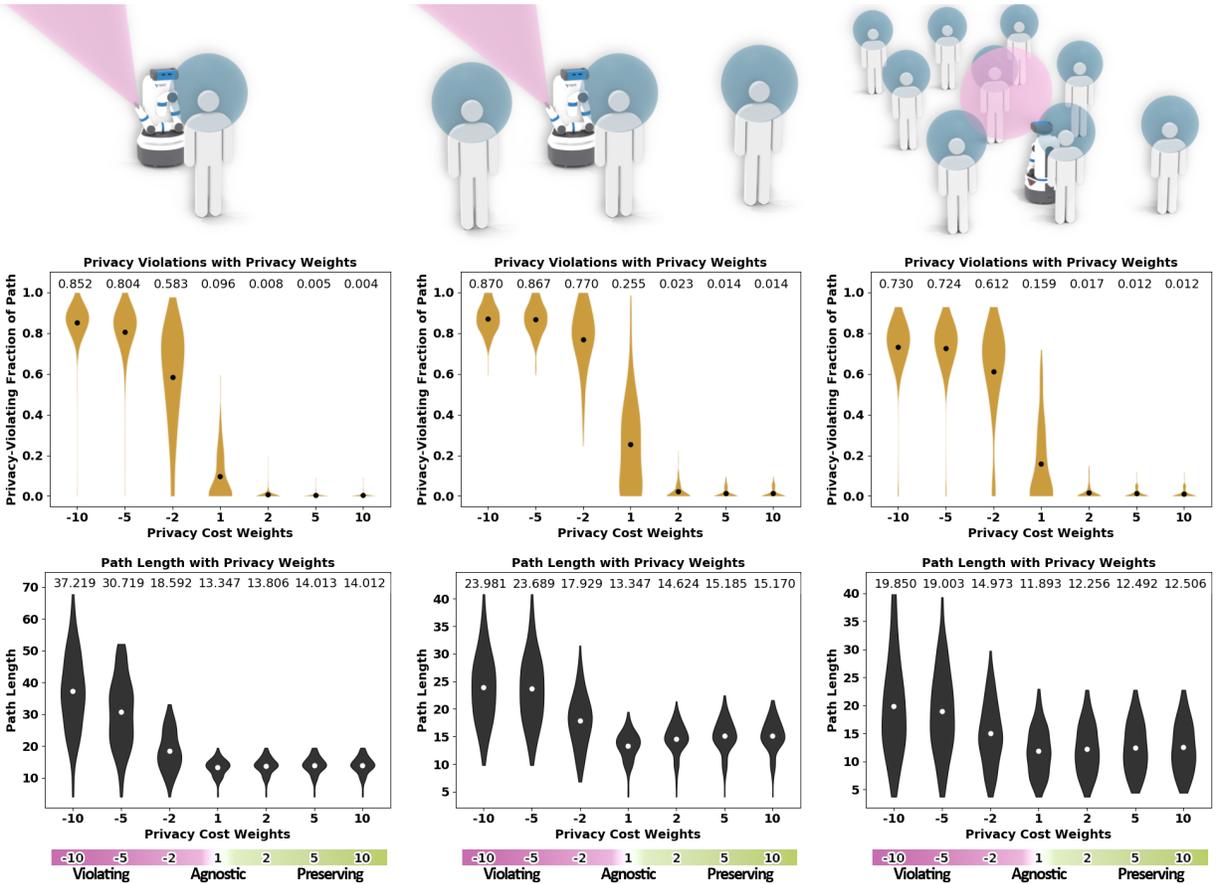


Fig. 3. (Top) A Fetch robot (with camera visibility cone in pink) controlling either the **arm+torso** with a hand camera (left, middle) or the **base+head** with a head camera (right). (From left to right) 1, 3, and 9 privacy regions (blue spheres) overlaid on humans in the workspace. Violin plots (with mean markers) for 100 random runs with the privacy violating fraction (middle) and path lengths (bottom) for different privacy weights. Mean values are given at the top.

its wrist (8-dim \mathcal{X}) and b) **base+head**: a controllable head camera (5-dim \mathcal{X}). A PRM* [79] is constructed and reused across the experiments. Each choice of privacy cost alters the weights on this roadmap. Given random valid starts and goals, uniform cost search reports a solution. The metrics reported for the resultant paths are a) **privacy violation fraction**, which is the fraction of the path where the sensing cone intersects with any of the privacy regions, and b) **path length**. w (Eqs. (1) and (2)) is chosen from $\{1, \pm 2, \pm 5, \pm 10\}$. The code was implemented with Robowflex [82] and OMPL [81].

C. Effects of Changing Privacy Cost (c_p)

Fig. 3 demonstrates the different behaviors obtained by changing c_p by tuning w . All the motions successfully connect the start and goal but differ in the interactions between the camera cone and privacy regions (Fig. 3 middle). Privacy-violating (negative) weights significantly increase the portion of the motions in which the camera *lingers* on the privacy regions. The privacy-violating solutions are also longer (Fig. 3 bottom) further increasing the data gathered. Note how an example violating motion (Fig. 4 left) ends up gathering data on all the regions. In contrast, the privacy-agnostic paths are shorter but are still liable to capture data on the privacy regions. The 3-region benchmark shows violations

along a quarter of the motions on average (Fig. 3 middle-middle). This motivates explicit reasoning about privacy preservation. Privacy-preserving paths immediately show the benefits of penalizing intersections between the cone and the privacy regions, with such violations dropping below 0.25% in all cases, while still maintaining relatively short paths. Note the privacy-preserving motion in Fig. 4 (right) completely avoids the privacy regions. Increasing the absolute magnitude of w strengthens the preserving or violating behavior.

A trade-off arises in the data between privacy violation and performance, both of which might be attributed economic value in the design of automation. Large negatively weighted privacy violations are associated with large path length increases. Interestingly, in certain cases ($w = -2$) high privacy violations (> 0.58) arise on average with only a small dip in performance (< 1.4 times the agnostic path length) in all the case studies. We highlight that double agent might be deliberately designed to **aggravate low-privacy, high-performance behaviors**. Additionally, trade-offs can arise when such data gathering might be necessary for detecting humans for safe operation and handling cases of dynamic or uncertain detections. Multiple sensors and targeted data gathering among multiple regions also poses risks. Though the studied cost function is *not* prescriptive and several alternate



Fig. 4. The visualization of different privacy-aware behavior: violating (left), agnostic (middle), and preserving (right) showing snapshots of the pink camera cone along motions solving the same problem. Even the agnostic path sweeps the camera cone over the right privacy region. The violating path makes the camera cone intersect with all the privacy regions while the preserving one avoids them altogether.

models exist, the chief takeaways remain unchanged.

Key Causes for Alarm: The observations from our simple simulated case studies already illustrate serious causes for concern. (a) With relatively straightforward alterations to the cost function, **drastic privacy-violating behavior was introduced**. (b) Significant privacy-violating behavior **might only introduce relatively small changes to the performance** (privacy-agnostic path length). (c) **Functionality to plug in custom cost function is readily available** in open-source planning libraries. (d) Privacy-preserving behavior **only emerges when deliberately included** in the cost function.

V. DISCUSSION

The work presented here has highlighted the privacy threats exposed by robotic double agents capable of autonomous planning and reasoning to move and sense in deliberate ways. This work contributes a thorough review of the interdisciplinary connections between privacy and motion planning, formulates the role of privacy in the motion planning problem, and showcases imminent privacy-violating threats using motivating simulated case studies and straightforward cost function modifications to motion planning. The work calls for a more consolidated investigation.

Human-Centric Factors: The human aspect is essential in the problem. The current simulated studies motivate the need for a deeper understanding of how humans perceive privacy-aware robotic behavior. The human-centric factors here are intrinsic to privacy, necessitating the investigations to be centered around the rights and protections of humans. There has been related human-robot interaction work that studies the anthropocentric principles to align robot motion to human expectations in intention-aware planning [83]–[85], the consideration of ethical or human-focused value alignment [86], and planning with legibility or human interpretation as an objective [87]–[89]. There has also been work on communication [90], understanding robot motion [91], [92], and parameterized social interactions [93] in navigation. Tradeoffs can exist between human-awareness in robots and privacy. The current study also raises questions about the context and expectations of privacy from the robot—an understanding by the human of an ongoing or imminent privacy violation.

Verification and Mitigation: This work admittedly raises more questions than answers with respect to the ways in which the threats posed by robotic planning can be identified and mitigated. While intentional deployment and recommended practices of privacy-preserving planning can achieve some headway, it does not resolve the issue of the bad actor or even a naïve one (who exposes threats that exist even in privacy-agnostic planning). There needs to be broader discussions among social scientists, ethicists, and policy-makers to inform rules, regulations, and deployments. The understanding of the human factors can also inform community stakeholders like coworkers or end-users. Privacy is also connected to vulnerabilities in open-source and general-purpose software. It is critical that the robotics community recognizes these imminent threats to step towards a future that avoids the worst of these threats.

Call to Action: We look into motion planning as one of the most fundamental capabilities of autonomous robots that can be abused for privacy violations. We demonstrate the imminent threats that exist on the near-horizon using technologies and solutions that exist today. The increased incidence of robots in our work and home will raise uncomfortable questions of how these robots are harvesting data and protecting privacy. A call to action is needed for the robotics community to reach out to other stakeholders to build towards a future with useful robots that are both capable and comply with fundamental human values, such as privacy.

REFERENCES

- [1] J. Wirtz, P. G. Patterson, W. H. Kunz, T. Gruber, V. N. Lu, S. Paluch, and A. Martins, “Brave new world: service robots in the Frontline,” *Journal of Service Management*, 2018.
- [2] M. Kangasniemi, S. Karki, N. Colley, and A. Voutilainen, “The use of robots and other automated devices in nurses’ work: an integrative review,” *International Journal of Nursing Practice*, vol. 25, no. 4, p. 12739, 2019.
- [3] D. Belanche, L. V. Casaló, C. Flavián, and J. Schepers, “Service robot implementation: a theoretical framework and research agenda,” *The Service Industries Journal*, vol. 40, no. 3-4, pp. 203–225, 2020.
- [4] M. A. K. Bahrin, M. F. Othman, N. H. N. Azli, and M. F. Talib, “Industry 4.0: a review on industrial automation and robotic,” *Jurnal Teknologi*, vol. 78, no. 6-13, 2016.
- [5] M. R. Calo, “Robots and privacy,” in *Machine Ethics and Robot Ethics*. Routledge, 2020, pp. 491–505.
- [6] K. Caine, S. Šabanovic, and M. Carter, “The effect of monitoring by cameras and robots on the privacy enhancing behaviors of older adults,” in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, 2012, pp. 343–350.

- [7] M. K. Lee, K. P. Tang, J. Forlizzi, and S. Kiesler, "Understanding users perception of privacy in human-robot interaction," in *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2011, pp. 181–182.
- [8] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, A. Y. Ng, et al., "Ros: an open-source robot operating system," in *ICRA Workshop on Open Source Software*, no. 3.2. Kobe, Japan, 2009, p. 5.
- [9] W. Hartzog, "Unfair and deceptive robots," *Md. L. Rev.*, vol. 74, p. 785, 2014.
- [10] A. Datta, M. C. Tschantz, and A. Datta, "Automated experiments on ad privacy settings: a tale of opacity, choice, and discrimination," *ArXiv Preprint ArXiv:1408.6491*, 2014.
- [11] S. Greengard, "Advertising gets personal," *Communications of the ACM*, vol. 55, no. 8, pp. 18–20, 2012.
- [12] T. Denning, C. Matuszek, K. Koscher, J. R. Smith, and T. Kohno, "A spotlight on security and privacy risks with future household robots: attacks and lessons," in *Proceedings of the 11th International Conference on Ubiquitous Computing*, 2009, pp. 105–114.
- [13] N. DeMarinis, S. Tellex, V. P. Kemerlis, G. Konidaris, and R. Fonseca, "Scanning the internet for ros: a view of security in robotics research," in *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, 2019, pp. 8514–8521.
- [14] E. Guo, "A Roomba recorded a woman on the toilet. how did screenshots end up on Facebook?" Dec. 2022.
- [15] S. Zuboff, "The age of surveillance capitalism: the fight for a human future at the new frontier of power," 2018.
- [16] S. Jana, A. Narayanan, and V. Shmatikov, "A scanner darkly: protecting user privacy from perceptual applications," in *2013 IEEE Symposium on Security and Privacy*. IEEE, 2013, pp. 349–363.
- [17] M. U. Kim, H. Lee, H. J. Yang, and M. S. Ryoo, "Privacy-preserving robot vision with anonymized faces by extreme low resolution," in *Proceedings of the IEEE/RSSJ International Conference on Intelligent Robots and Systems*. IEEE, 2019, pp. 462–467.
- [18] D. J. Butler, J. Huang, F. Roesner, and M. Cakmak, "The privacy-utility tradeoff for remotely Teleoperated robots," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 2015, pp. 27–34.
- [19] K. Martin and K. N. Plataniotis, "Privacy protected surveillance using secure visual object coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 8, pp. 1152–1162, 2008.
- [20] J. DeCew, "Privacy," in *The Stanford Encyclopedia of Philosophy*, spring 2018 ed., E. N. Zalta, Ed. Metaphysics Research Lab, Stanford University, 2018.
- [21] P. Voigt and A. Von dem Bussche, "The eu general data protection regulation (Gdpr)," *A Practical Guide, 1St Ed., Cham: Springer International Publishing*, vol. 10, no. 3152676, pp. 10–5555, 2017.
- [22] C. de l'Europe et al., *Convention for the protection of individuals with regard to automatic processing of personal data*. Council of Europe, 1981, vol. 108.
- [23] B. Dieber, S. Kacianka, S. Rass, and P. Schartner, "Application-level security for ros-based applications," in *Proceedings of the IEEE/RSSJ International Conference on Intelligent Robots and Systems*. IEEE, 2016, pp. 4477–4482.
- [24] J. McClean, C. Stull, C. Farrar, and D. Mascarenas, "A preliminary Cyber-physical security assessment of the robot operating system (ros)," in *Unmanned Systems Technology Xv*, vol. 8741. SPIE, 2013, pp. 341–348.
- [25] B. Dieber, B. Breiling, S. Taurer, S. Kacianka, S. Rass, and P. Schartner, "Security for the robot operating system," *Robotics and Autonomous Systems*, vol. 98, pp. 192–203, 2017.
- [26] R. White, D. Christensen, I. Henrik, D. Quigley, et al., "SROS: securing ros over the wire, in the graph, and through the kernel," *ArXiv Preprint ArXiv:1611.07060*, 2016.
- [27] A. B. Nichols, "Difficult choices: the ethical dimensions of engineering," *Water Environment & Technology*, vol. 2, no. 2, pp. 60–67, 1990.
- [28] J. Van den Hoven, G.-J. Lokhorst, and I. Van de Poel, "Engineering and the problem of moral overload," *Science and Engineering Ethics*, vol. 18, no. 1, pp. 143–155, 2012.
- [29] C. Lutz and A. Tamò, "RoboCode-ethicists: privacy-friendly robots, an ethical responsibility of engineers?" in *Proceedings of the ACM Web Science Conference*, 2015, pp. 1–12.
- [30] M. U. Hassan, M. H. Rehmani, and J. Chen, "Differential privacy techniques for Cyber physical systems: a survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 746–789, 2019.
- [31] J. Giraldo, E. Sarker, A. A. Cardenas, M. Maniatakos, and M. Kantarcioglu, "Security and privacy in Cyber-physical systems: a survey of surveys," *IEEE Design & Test*, vol. 34, no. 4, pp. 7–17, 2017.
- [32] R. H. Weber, "Internet of things—new security and privacy challenges," *Computer Law & Security Review*, vol. 26, no. 1, pp. 23–30, 2010.
- [33] M. E. Kaminski, M. Rueben, W. D. Smart, and C. M. Grimm, "Averting robot eyes," *Md. L. Rev.*, vol. 76, p. 983, 2016.
- [34] B. A. Mills, "Alexa, i want the truth!": a prosecutor's guide to the collection and use of evidence from virtual assistants," *Army Law*, p. 48, 2021.
- [35] M. Harrigan, "Privacy versus justice: amazon's first amendment battle in the cloud," *W. St. UL Rev.*, vol. 45, p. 91, 2017.
- [36] E. McReynolds, S. Hubbard, T. Lau, A. Saraf, M. Cakmak, and F. Roesner, "Toys that listen: a study of parents, children, and internet-connected toys," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 5197–5207.
- [37] A. Das, M. Degeling, X. Wang, J. Wang, N. Sadeh, and M. Satyanarayanan, "Assisting users in a world full of cameras: a privacy-aware infrastructure for computer vision applications," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2017, pp. 1387–1396.
- [38] J. Schiff, M. Meingast, D. K. Mulligan, S. Sastry, and K. Goldberg, "Respectful cameras: detecting visual markers in real-time to address privacy concerns," in *Protecting Privacy in Video Surveillance*. Springer, 2009, pp. 65–89.
- [39] N. Raval, A. Srivastava, K. Lebeck, L. Cox, and A. Machanavajjhala, "Markit: privacy markers for protecting visual secrets," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, 2014, pp. 1289–1295.
- [40] F. E. Fernandes, G. Yang, H. M. Do, and W. Sheng, "Detection of privacy-sensitive situations for social robots in smart homes," in *Proceedings of the IEEE International Conference on Automation Science and Engineering*. IEEE, 2016, pp. 727–732.
- [41] M. Rueben, F. J. Bernieri, C. M. Grimm, and W. D. Smart, "Evaluation of physical marker interfaces for protecting visual privacy from mobile robots," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2016, pp. 787–794.
- [42] F. Loula, S. Prasad, K. Harber, and M. Shiffrar, "Recognizing people from their movement," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 31, no. 1, p. 210, 2005.
- [43] E. Glikson and A. W. Woolley, "Human trust in artificial intelligence: review of empirical research," *Academy of Management Annals*, vol. 14, no. 2, pp. 627–660, 2020.
- [44] M. Rueben, M. J. Matarić, E. Rothberg, and M. Tang, "Estimating and influencing user mental models of a robot's perceptual capabilities: initial development and pilot study," in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 418–420.
- [45] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and Autonomous Systems*, vol. 57, no. 5, pp. 469–483, 2009.
- [46] H. Ravichandar, A. S. Polydoros, S. Chernova, and A. Billard, "Recent advances in robot learning from demonstration," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, pp. 297–330, 2020.
- [47] R. Shokri and V. Shmatikov, "Privacy-preserving deep learning," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, 2015, pp. 1310–1321.
- [48] M. Al-Rubaie and J. M. Chang, "Privacy-preserving machine learning: threats and solutions," *IEEE Security & Privacy*, vol. 17, no. 2, pp. 49–58, 2019.
- [49] C. Dwork, "Differential privacy: a survey of results," in *International Conference on Theory and Applications of Models of Computation*. Springer, 2008, pp. 1–19.
- [50] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 308–318.
- [51] S. Han, G. J. Pappas, et al., "Privacy in control and dynamical systems," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, no. 1, pp. 309–332, 2018.
- [52] S. Goldwasser, M. P. Kim, V. Vaikuntanathan, and O. Zamir, "Planting undetectable backdoors in machine learning models," in *2022 IEEE*

- 63rd Annual Symposium on Foundations of Computer Science (FOCS). IEEE, 2022, pp. 931–942.
- [53] A. Jobin, M. Ienca, and E. Vayena, “The global landscape of AI ethics guidelines,” *Nature Machine Intelligence*, vol. 1, no. 9, pp. 389–399, 2019.
- [54] M. Rueben, A. M. Aroyo, C. Lutz, J. Schmölz, P. Van Cleynenbreugel, A. Corti, S. Agrawal, and W. D. Smart, “Themes and research directions in privacy-sensitive robotics,” in *2018 IEEE Workshop on Advanced Robotics and Its Social Impacts (ARSO)*. IEEE, 2018, pp. 77–84.
- [55] S. Eick and A. I. Antón, “Enhancing privacy in robotics via judicious sensor selection,” in *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, 2020, pp. 7156–7165.
- [56] C. Lutz, M. Schöttler, and C. P. Hoffmann, “The privacy implications of social robots: scoping review and expert interviews,” *Mobile Media & Communication*, vol. 7, no. 3, pp. 412–434, 2019.
- [57] E. Fosch-Villaronga, C. Lutz, and A. Tamò-Larrieux, “Gathering expert opinions for social robots’ ethical, legal, and societal concerns: findings from four international workshops,” *International Journal of Social Robotics*, vol. 12, no. 2, pp. 441–458, 2020.
- [58] S. Chen, Y. Li, and N. M. Kwok, “Active vision in robotic systems: a survey of recent developments,” *The International Journal of Robotics Research*, vol. 30, no. 11, pp. 1343–1377, 2011.
- [59] B. R. Abidi, N. R. Aragam, Y. Yao, and M. A. Abidi, “Survey and analysis of multimodal sensor planning and integration for wide area surveillance,” *ACM Computing Surveys (CSUR)*, vol. 41, no. 1, pp. 1–36, 2009.
- [60] K. Shubina and J. K. Tsotsos, “Visual search for an object in a 3d environment using a mobile robot,” *Computer Vision and Image Understanding*, vol. 114, no. 5, pp. 535–547, 2010.
- [61] A. Adu-Bredu, N. Devraj, P.-H. Lin, Z. Zeng, and O. C. Jenkins, “Probabilistic inference in planning for partially observable long horizon problems,” in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2021, pp. 3154–3161.
- [62] H. Rahmani, D. A. Shell, and J. M. O’Kane, “Planning to chronicle,” in *Proceedings of the Workshop on the Algorithmic Foundations of Robotics*. Springer, 2020, pp. 277–293.
- [63] D. A. Shell, L. Huang, A. T. Becker, and J. M. O’Kane, “Planning coordinated event observation for structured narratives,” in *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, 2019, pp. 7632–7638.
- [64] D. Rakita, B. Mutlu, and M. Gleicher, “An autonomous dynamic camera method for effective remote Teleoperation,” in *2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2018, pp. 325–333.
- [65] C. Armbrust, S. A. Mehdi, M. Reichardt, J. Koch, and K. Berns, “Using an autonomous robot to maintain privacy in assistive environments,” *Security and Communication Networks*, vol. 4, no. 11, pp. 1275–1293, 2011.
- [66] A. Tsiamis, A. B. Alexandru, and G. J. Pappas, “Motion planning with secrecy,” in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 784–791.
- [67] L. Li, A. Bayuelo, L. Bobadilla, T. Alam, and D. A. Shell, “Coordinated multi-robot planning while preserving individual privacy,” in *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, 2019, pp. 2188–2194.
- [68] H. Zheng, J. Panerati, G. Beltrame, and A. Prorok, “An adversarial approach to private flocking in mobile robot teams,” *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1009–1016, 2020.
- [69] M. Nieuwenhuisen and S. Behnke, “Search-based 3d planning and trajectory optimization for safe micro aerial vehicle flight under sensor visibility constraints,” in *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, 2019, pp. 9123–9129.
- [70] Y. Luo, Y. Yu, Z. Jin, Y. Li, Z. Ding, Y. Zhou, and Y. Liu, “Privacy-aware UAV flights through self-configuring motion planning,” in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2020, pp. 1169–1175.
- [71] Y. Pan, S. Li, J. L. Chang, Y. Yan, S. Xu, Y. An, and T. Zhu, “An unmanned aerial vehicle navigation mechanism with preserving privacy,” in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*. IEEE, 2019, pp. 1–6.
- [72] H. Kim, J. Ben-Othman, and L. Mokdad, “UDiPP: a framework for differential privacy preserving movements of unmanned aerial vehicles in smart cities,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3933–3943, 2019.
- [73] P. McBride, “Beyond orwell: the application of unmanned aircraft systems in domestic surveillance operations,” *J. Air L. & Com.*, vol. 74, p. 627, 2009.
- [74] R. L. Finn and D. Wright, “Unmanned aircraft systems: surveillance, ethics and privacy in civil applications,” *Computer Law & Security Review*, vol. 28, no. 2, pp. 184–194, 2012.
- [75] P. Blank, S. Kirrane, and S. Spiekermann, “Privacy-aware restricted areas for unmanned aerial systems,” *IEEE Security & Privacy*, vol. 16, no. 2, pp. 70–79, 2018.
- [76] B. S. Stewart and C. C. White III, “Multiobjective a,” *Journal of the ACM (JACM)*, vol. 38, no. 4, pp. 775–814, 1991.
- [77] B. Goldin and O. Salzman, “Approximate bi-criteria search by efficient representation of subsets of the Pareto-optimal frontier,” in *Proceedings of the International Conference on Automated Planning and Scheduling*, vol. 31, 2021, pp. 149–158.
- [78] D. Devaurs, T. Siméon, and J. Cortés, “Enhancing the transition-based RRT to deal with complex cost spaces,” in *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 4120–4125.
- [79] S. Karaman and E. Frazzoli, “Sampling-based algorithms for optimal motion planning,” *The International Journal of Robotics Research*, vol. 30, no. 7, pp. 846–894, 2011.
- [80] L. E. Kavraki, P. Švestka, J.-C. Latombe, and M. Overmars, “Probabilistic Roadmaps for path planning in high dimensional configuration spaces,” *IEEE Transactions on Robotics and Automation*, vol. 12, no. 4, pp. 566–580, 1996.
- [81] I. A. Sucan, M. Moll, and L. E. Kavraki, “The open motion planning library,” *IEEE Robotics and Automation Magazine*, vol. 19, no. 4, pp. 72–82, 2012.
- [82] Z. Kingston and L. E. Kavraki, “Robowflex: robot motion planning with MoveIt made easy,” in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2022.
- [83] T. Bandyopadhyay, K. S. Won, E. Frazzoli, D. Hsu, W. S. Lee, and D. Rus, “Intention-aware motion planning,” in *Algorithmic Foundations of Robotics X*. Springer, 2013, pp. 475–491.
- [84] H. Bai, S. Cai, N. Ye, D. Hsu, and W. S. Lee, “Intention-aware online POMDP planning for autonomous driving in a crowd,” in *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, 2015, pp. 454–460.
- [85] M. Chen, S. Nikolaidis, H. Soh, D. Hsu, and S. Srinivasa, “Planning with trust for human-robot collaboration,” in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 2018, pp. 307–315.
- [86] D. Hadfield-Menell, S. J. Russell, P. Abbeel, and A. Dragan, “Cooperative inverse reinforcement learning,” *Advances in Neural Information Processing Systems*, vol. 29, 2016.
- [87] C. Lichtenthaler and A. Kirsch, “Towards legible robot navigation-how to increase the intend expressiveness of robot navigation behavior,” in *International Conference on Social Robotics-Workshop Embodied Communication of Goals and Intentions*, 2013.
- [88] A. D. Dragan, K. C. Lee, and S. S. Srinivasa, “Legibility and predictability of robot motion,” in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2013, pp. 301–308.
- [89] A. D. Dragan, S. Bauman, J. Forlizzi, and S. S. Srinivasa, “Effects of robot motion on human-robot collaboration,” in *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2015, pp. 51–58.
- [90] S. H. Huang, D. Held, P. Abbeel, and A. D. Dragan, “Enabling robots to communicate their objectives,” *Autonomous Robots*, vol. 43, no. 2, pp. 309–326, 2019.
- [91] M. Beetz, F. Stulp, P. Esden-Tempski, A. Fedrizzi, U. Klank, I. Kresse, A. Maldonado, and F. Ruiz, “Generality and legibility in mobile manipulation,” *Autonomous Robots*, vol. 28, no. 1, pp. 21–44, 2010.
- [92] A. Dragan and S. Srinivasa, “Familiarization to robot motion,” in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, 2014, pp. 366–373.
- [93] D. V. Lu, D. B. Allan, and W. D. Smart, “Tuning cost functions for social navigation,” in *International Conference on Social Robotics*. Springer, 2013, pp. 442–451.