

1 INTRODUCTION

People have grown increasingly frustrated with the decisions that algorithmic systems make over their lives. These decisions can have weighty consequences: they determine whether we’re excluded from social environments [? ?]; they decide whether we should be paid for our work [?]; they influence whether we’re sent to jail or released on bail [?]. Despite the importance of getting these decisions right, these algorithmic and machine learning systems make surprising and frustrating errors: they ban good actors from social environments and leave up toxic content [? ?]; they disproportionately make bail inaccessible for people of color [?]; they engage in wage theft for honest workers [?]. Across diverse applications of algorithmic systems, one aspect seems to come through: surprise and — quite often — frustration with these systems over the decisions they make.

Researchers have approached these problems in algorithmic systems from a number of perspectives: some have interrogated the hegemonic influence that these systems have over their users [?]; others have documented and called attention to the unfair and opaque decisions these systems sometimes make [?]; others still have audited these algorithms via the criteria of fairness, accountability and transparency [?]. But even if we could answer the questions at the hearts of these research agendas — finding clear guiding principles, incorporating the needs of myriad stakeholders, and ensuring fairness — these algorithms will always face cases that are *at the margin*: outside the situations seen in their training data. For example, even a value-sensitive [?], fair, and transparent content moderation algorithm will likely make errors with high confidence when classifying content with a new slur that it has never seen before. And it is exactly at these moments that the algorithm has to generalize: to “fill in the gaps” between the policy implied by training data and a new case the likes of which it has never seen before.

“Filling in the gaps” is not a problem unique to algorithms, however: it has been the chief concern for decades of what bureaucratic theorists call *street-level bureaucracies* [?]. A street-level bureaucracy is the layer of a bureaucracy that directly interacts with people. They are responsible for making decisions “on the street”, filling in the gaps between legislated policies and the situations in front of them. Street-level bureaucrats are police officers, judges, teachers, customer service agents, and others with whom members of the public interact frequently, and who make everyday decisions that affect our lives throughout the day. In all of these cases, street-level bureaucrats make important decisions about both familiar and new cases that require them to translate an official policy into the situations they face. A police officer must choose whether to issue a warning or a traffic citation; a judge decides whether to allow a defendant to pay bail or to have them remanded to jail; a teacher determines whether to waive a course’s prerequisites for a student. These decisions may have involve nuance extenuating circumstances in every case, making it all but impossible to prescribe the right response for all situations.

Street-level bureaucracies are important because the decisions they make are the manifestation of the power of the institution, and are *effectively* policy. Regardless of what’s explicitly legislated or prescribed, policy is effected by the street-level bureaucrat’s execution of their duties. Police officers choose whether to issue citations to people speeding only slightly over the speed limit; judges may make every effort to make bail accessible given a defendant’s circumstances; teachers may allow some flexibility in course prerequisites. In each case, these street-level decisions become expectations of the system writ large. When these effective policies are biased (e.g., [?]), it prompts broad critiques and reviews of the policy, organization, or system.

In this paper, we draw the analogy to pose people’s interactions with algorithmic aspects of sociotechnical systems as interactions with what we term *street-level algorithms*. Street-level algorithms are algorithmic systems that directly interact with and make decisions about people in a sociotechnical system. They make on-the-ground decisions about human lives and welfare, filling in the gaps between platform policy and implementation, and represent the algorithmic layer that mediates interaction between humans and complex computational systems.

Our central claim is that street-level algorithms make frustrating decisions in many situations where street-level bureaucrats do not, because street-level bureaucrats can reflexively refine the contours of their decision boundary *before* making a decision on a novel or marginal case, but street-level algorithms at best refine these contours only *after* they make a decision. Our focus here is on cases at the margin: those representing marginal or underrepresented groups, or others creating novel situations not seen often or at all in the training data. When street-level bureaucrats encounter a novel or marginal case, they use that case to reflexively refine their interpretation of the policies. When street-level algorithms encounter a novel or marginal case, they execute their pre-trained classification boundary, potentially with erroneously high confidence [?]. For a bureaucrat, but not an algorithm, policy execution is also a reflexive action. For an algorithm, but not for a bureaucrat, reflexivity must wait until after the system receives reward feedback or additional training data. The result is that street-level algorithms unintentionally make nonsensical decisions, never revisiting the decision or the motivating rationale it until it has prompted human review.

We will examine a series of case studies through the lens of street-level algorithms. First, we will examine why YouTube's demonetization algorithms deny payment to marginalized groups such as LGBTQ content creators. Second, we will look at judicial bail-recommendation systems and interrogate their biases — for example, disproportionately recommending jail for people of color. Finally, we will analyze the algorithmic management of crowd workers, discussing problems that arise when algorithmic systems evaluate the quality of workers who must themselves make interpretations about underspecified tasks. In all of these cases, we'll illustrate the difference between policy explication and policy execution, or in some other sense, the difference between training data and testing data.

We will discuss ways that designers of sociotechnical systems can mitigate the damage that street-level algorithms do in marginal cases. Prior work has called for routes of appeal and recourse when a person is misclassified by an algorithmic system [?]. However, in practice, many appeals are quickly denied. The lens of street-level algorithms allows us to refine this claim: people might compare their case to similar cases seen by the algorithm, arguing specifically how the algorithm's decision boundary misunderstood their case relative to its prior experiences. For example, LGBTQ content creators on YouTube could appeal on the grounds that the examples provided by the algorithm were all sexually explicit, and that their videos were categorically different: explicitly about sexuality, rather than sexually explicit. If the algorithm operated on a decision boundary and was not aware of the marginality or novelty of the situation, this appeal makes clearer how and why the algorithm should be retrained.

Many discussions of the invisible systems in human-computer interaction might also be considered interrogations of street-level algorithms. When we talk about systems that decide which social media posts we see [?], when we find ourselves in conflict with Wikipedia bots [?], and when we enforce screen time with our children [?], we are creating, debating, and designing around street-level algorithms. Our hope is that the theoretical lens introduced in this paper can help make sense of these previously disconnected phenomena.