Good afternoon, Chair Vacca and members of the Committee on Technology. My name is Julia Powles, and I am a Research Fellow in the Digital Life Initiative at Cornell Tech, New York City’s bold new interdisciplinary research and technology campus at Roosevelt Island. I am joined in providing this testimony by two of my Cornell Tech colleagues, Helen Nissenbaum, Professor of Information Science and Director of the Digital Life Initiative, and Thomas Ristenpart, Associate Professor of Computer Science. Our work over many years speaks to various angles of this afternoon’s hearing, but perhaps most pertinently, this June, Professors Nissenbaum and Ristenpart launched a major, multi-year, NSF-funded research project led by Cornell Tech, Carnegie Mellon University, and the International Computer Science Institute, Berkeley, to investigate threats to privacy and fairness in automated decision-making systems and, in particular, to develop mechanisms for accountable information use in such systems.¹

The most important work that a Bill in the area of automated systems can do is to build accountability—both the accountability of vendors of these systems to the City, and the accountability of the City and its agencies to the people of New York. This Bill is an ambitious attempt to seek accountability through transparency. It attempts this by two means: first, a requirement of source code disclosure; and second, by mediating what is known in the field as “black-box testing”—a mechanism for testing inputs, generating outputs, and deriving insights. Before elaborating further, we would like to add to the chorus of those who have expressed appreciation to Council Member Vacca for bringing forward this proposal, as well as to the Committee on Technology for its important efforts to bring greater transparency—and, potentially, accountability—to automated systems that profoundly affect people’s lives and life chances. This legislative direction is both exciting and essential. It offers the City Administration the opportunity to be a real bellwether, not only of good governance and government, but of innovation of the truest kind, stimulating better technologies and better civic engagement.

In this testimony, we suggest ways to make the Bill more effective in realizing its ambitions, and offer concrete recommendations to improve it.

A Bill like this has the potential to address several stark gaps in our regulatory landscape. When data is fed into a computer system and used to allocate public services, penalties, or policing, people deserve to know that the system is functioning in accordance with the City’s aims and values. That it is not arbitrary, unfair, or incorrect. That it does not amplify inequality. This means being able to find out what data is used, how it is processed, and what else is taken into

consideration in decision-making, both in general and in individual cases. There should be opportunities to test and contest the input, processing, and output. In other words, there is a need for accountable systems, including clear processes for calling to account responsible parties (those designing, procuring, or using systems), if there is cause for complaint, or even suspicion that systems under consideration or in use are failing to meet aims and values.

This Bill makes important advances, but in order to meet these critical ends, there are dimensions it does not yet address in its current formulation. For example, it gives no view onto the data that is being used by an automated system, nor how decisions are made in general or individual cases. Some capacity to test and understand systems is offered, but if what is seen in the source code (if it is even provided) or through black-box testing is unsatisfactory, there is no direct mechanism for contestation and bringing responsible parties to account. In other words, ambitious and important though this Bill is, it relies on a degree of accountability emerging as a by-product of a very particular, and potentially limited, kind of public transparency. It is well within the capacity of New York City to tackle the targets of both transparency and accountability much more directly.

A primary source of these limitations comes from the location of the Bill and the provisions that surround it. The section of the Administrative Code where the provisions are proposed to be inserted concerns “Open Data.” This fundamentally affects the nature and impact of the Bill as it is currently drafted. It means, crucially, that, according to section 23-504(c) of the Code, the Bill gives rise to no actionable rights, either for individuals or against an agency. Section 23-504(a) makes clear that data is provided to the public only “for informational purposes,” with section 23-504(b) clarifying that there are no guarantees as to “completeness, accuracy, content, or fitness for use.” Further, the Bill’s placement within the Open Data provisions also means that, following the logic of sections 23-501(g), any proprietary claims and intellectual property assertions in relation to the code and systems, no matter how broad or baseless, will readily thwart the intent to provide transparency.

It may be that locating these provisions in the Open Data provisions is regarded as optimal for other reasons, such as uniform Council support for the City’s commitment to open public processes, but the legislative context should be given further and careful consideration. If it is resolved that the present location remains the most desirable among a range of legislative options, the Bill should be elaborated, and the applicability or otherwise of the remainder of the Open data provisions should be explicitly addressed, particularly those concerning private rights of action, liability of agencies, and the tension between disclosure of the source code and operations of automated systems and proprietary interests.

Turning to the aspect of the Bill that concerns black-box testing, the requirement as proposed is likely to be administratively burdensome on agencies, to the point of potential impracticality, given the dynamism of automated systems, and the fact that effective black-box testing in the public interest can require thousands of queries (or more), depending on the context; a prospect that is likely to be highly constrained if every query must be mediated through an agency request.

Collectively, these realities mean that the proposed Bill as it stands has limited purchase on the target of accountability, and that there are significant obstacles to it being the strong instrument of algorithmic transparency that it could be, though it makes important strides in this direction.
None of these weaknesses are fatal, and the necessary leverage is well within the remit and capacity of the New York City Administration to address. Council Members committed to the important cause of improving the transparency and accountability of automated systems have a number of options available to them to help realize the objectives behind this Bill. We have eight recommendations.

First, given the limitations of the Open Data provisions, consider other parts of the Code where these provisions might be better located.

Second, ensure that automated systems used for the provision of public services, penalties, and policing disclose all data sources that they incorporate, as well as additional parameters about the data selected for training, model choices, excluded data, and other standardized requirements for best practice disclosure for interpretability and accountability.

Third, ensure that intellectual property and other proprietary rights’ assertions cannot be used to defeat algorithmic transparency requirements. Explore qualified transparency if public transparency is not possible.

Fourth, develop mechanisms to tie transparency requirements more strongly to enforcement, such as through making City funding of agencies conditional upon meeting certain explicit standards of algorithmic disclosure and interpretability, audited by independent expert assessors.

Fifth, establish private rights of action for systems that are found to be unsatisfactory.

Sixth, in consultation with experts, establish benchmarks for best practices for source code disclosure. Establish how frequently code updates should be notified, as well as any limitations on disclosure, publication, and retention.

Seventh, in consultation with experts, establish how black-box testing requirements are going to be managed at a practical level. Provide examples for how outputs of user-submitted tests will be provided to users. Ensure that third party testing in the public interest, sometimes requiring thousands of queries, can be managed without becoming burdensome on agencies.

Eighth, institute a City-wide practice that when agencies engage vendors of automated systems, any data sources continue to be managed in the public interest. This entails non-exclusive data use, transparency about integration with other data sources, and ongoing public stewardship over the data—whether in raw, cleaned, catalogued, or systematized form.

These are our recommendations. Thank you for your time, and we applaud you again for this bold and inspiring legislative effort. Cornell Tech and the Digital Life Initiative are dedicated to the development and deployment of technologies in the public interest, and we are committed to being a partner to the City in this essential endeavor.