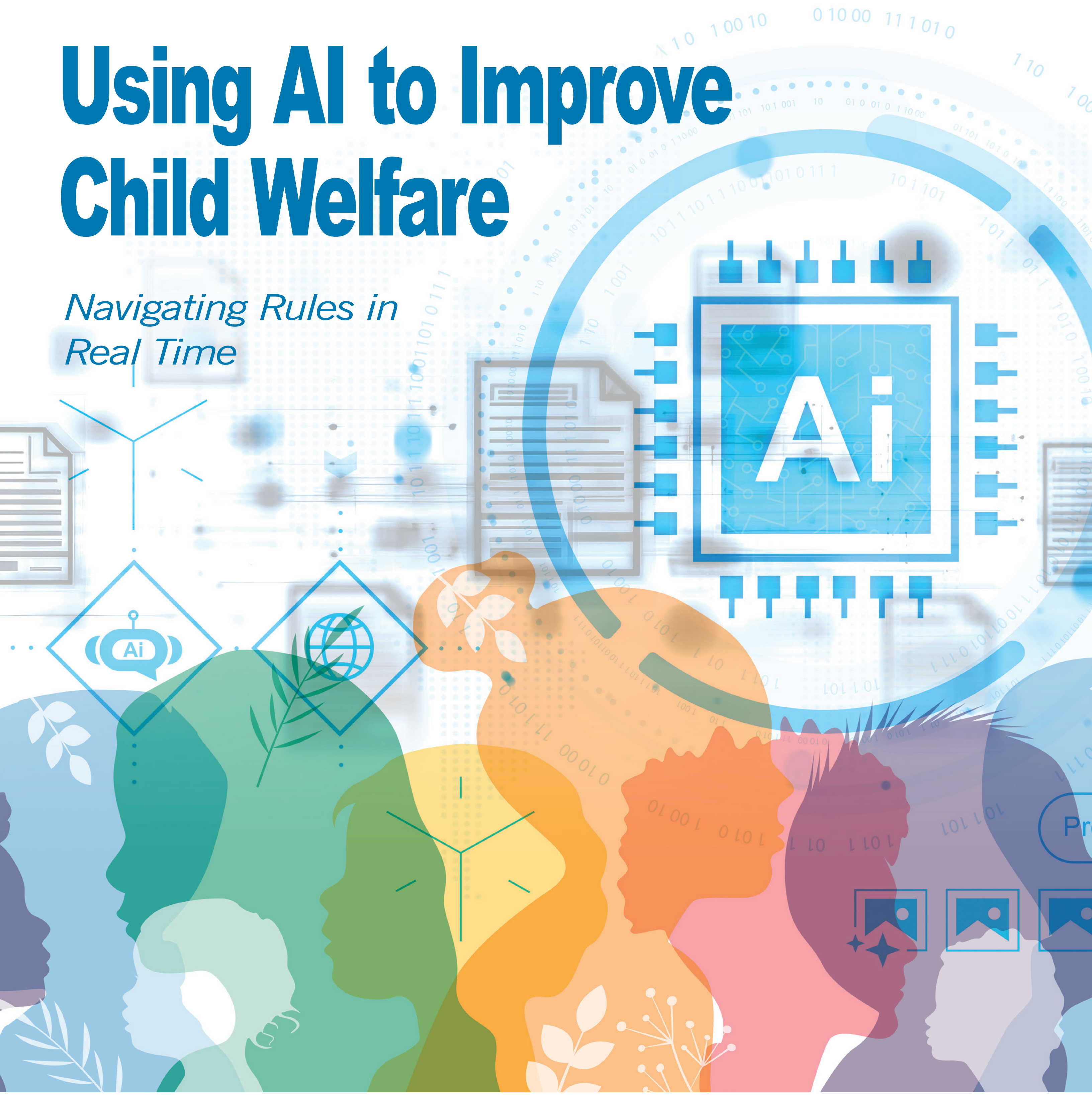
Using AI to Improve Child Welfare

Navigating Rules in Real Time



David R. Schwartz, MSW

SCHOOL OF SOCIAL WORK
CHILD & ADOLESCENT DATA LAB
UNIVERSITY OF MICHIGAN

IBM Center for The Business of Government

Using AI to Improve Child Welfare:
Navigating Rules in Real Time

David R. Schwartz, MSW
AI Policy Practitioner & Social Work Researcher

APRIL 2026

IBM Center for
**The Business
of Government**

Table of Contents

“Ultimately, the goal is to free workers from the administrative morass so they can focus on what only humans can do—build relationships, read body language, earn trust, assess danger, keep children safe.”

Foreword

Daniel J. Chenok

Brian E. Perron

Joseph P. Ryan

Leah Pewitt

Across the country, child welfare agencies carry out a vital public mission amid growing complexity—supporting increasing caseloads, navigating evolving policy frameworks, and continually developing a workforce with skills to make consequential decisions for children and families.

At the same time, these agencies must document every action, comply with mandates from multiple levels of government, and deliver high-stakes outcomes. The gap between what the child welfare workforce has the responsibility to do, relative to what existing systems enable them to do, has widened in recent years. States have begun to adopt AI as a means to close that gap.

This report, *Using AI to Improve Child Welfare: Navigating Rules in Real Time* by David R. Schwartz, published by the IBM Center for The Business of Government in collaboration with the University of Michigan School of Social Work, addresses this issue directly. Drawing on insights from a national roundtable of state and local child welfare leaders, researchers, and practitioners, the author explores how responsible use of AI can support frontline social workers, supervisors, and agency leaders—by reducing administrative burden, improving access to policy and case information, and strengthening professional judgment and accountability.

This report reflects a growing recognition across government that successful innovation in human services must support the frontline workforce that serves families and children in need. The report presents insights about what AI can do, cautions about ethical risks, and discusses how technology can serve people.

The report makes clear that the promise of AI in child welfare lies not in automation of decisions about child safety, but rather in removing administrative burdens that have made this work increasingly challenging. The AI tools described in this report focus on answering policy questions in real-time, synthesizing complex case histories, assisting with documentation, and supporting training—all while keeping humans in the loop. These advances are grounded in early, practical implementations already underway in multiple states.

This report integrates historical perspective, applied research, and practitioner experience to frame a set of principles for responsible adoption. These principles can support relief rather than replacement, transparency over opacity, low-risk use cases before high-stakes ones, and governance structures that support emerging technology. From this foundation emerge clear lessons that can shape key imperatives, including local control of sensitive data and collaboration across agencies and disciplines.

Importantly, the author underscores that outcomes in child welfare cannot emerge only from automation, but require informed design, governance, evaluation, and stewardship. AI systems can enhance progress toward these goals, but only with sufficient oversight, feedback loops, and continuous learning. When implemented thoughtfully, AI can enable the frontline workforce to spend less time navigating systems and processes and more time doing what humans do best: assess risk, support families, build trust, and keep children safe.

This report builds on the IBM Center's longstanding body of work examining how effective use of technology, data, and management practices can strengthen government performance, including *Responsible AI for Public Evaluation, AI in State Government, GenAI and the Future of Government Work*, and *AI and the Modern Tax Agency*.

As governments continue to explore the responsible use of AI for service delivery, *Using AI to Improve Child Welfare* offers a grounded and pragmatic path forward. By showing a path to innovation that supports practitioners and the needs of vulnerable populations, this report contributes findings and recommendations to ensure that technology enhances public value at the heart of child welfare.

Daniel J. Chenok
Executive Director
IBM Center for
The Business of Government
chenokd@us.ibm.com

Brian E. Perron
School of Social Work,
University of Michigan
beperron@umich.edu

Joseph P. Ryan
School of Social Work,
University of Michigan
joryan@umich.edu

Leah Pewitt
Associate Partner, Child Welfare
Practice Lead, IBM
ldienge@us.ibm.com

Introduction: The Forest and the Trees

Picture a forest where the youngest saplings bear the heaviest fruit. Every year, new trees are planted at the edge—and every year, those trees are expected to hold up an entire canopy. This is akin to the reality of child welfare work in America.

When a preschool teacher calls in a concern, when a neighbor reports something troubling, when a child shows up at school with bruises that don't quite add up—someone has to knock on a family's door. Someone has to walk into a stranger's home, assess a potentially chaotic situation in minutes, and make a high-stakes decision. This impossible calculus faces caseworkers daily.

Turnover in child welfare agencies is significant. Some climb quickly into supervisory roles, leaving the frontlines behind. Others simply leave—seeking work where mistakes won't end up on the local news, or where the impact of a wrong judgment won't follow them home at night. And so, the cycle continues: new workers arrive, get trained, become overwhelmed, and often either move up or move out. The frontlines of child welfare agencies remain perpetually staffed by people who are learning on the job to serve a high-risk population.

The State of Florida, one of the largest child welfare systems in the United States, has acknowledged the high-stress nature of this job, publishing that its turnover rate for Child Protective Investigators was 64 percent in FY 2023-2024.[1] In Arizona, the annual turnover rate among frontline social workers is reported to be approximately 35 percent; however, this figure may be modest, given that it encompasses all frontline roles rather than the high-stress position of Child Protective Services Investigator.

What if these workers had an experienced partner from day one—one that never gets overwhelmed, never gets pulled into court, and always knows the latest policy?

1. Navigating the Policy Maze

Consider the example of a new caseworker in week three, facing their first decision to remove a child from a home. They sense the child may be in danger, but they must simultaneously navigate a maze of policies: federal guidelines, state statutes, agency protocols, court requirements, and tribal sovereignty laws if the family has Native heritage. One wrong step—removing a child without proper legal grounds, or failing to remove when they should—could shatter a family or affect a child's life. Their training manual is 300 pages. The policy database contains 47 different documents. Their supervisor is in court for the morning.

What if, instead of frantically searching through binders or hoping they remember the right protocol from a training session two months ago, they could simply ask: What are the legal requirements for emergency removal in this situation? . . . and then they receive an instant, accurate answer grounded in current policy—not to make the decision for them, but to ensure they are operating within legal and policy boundaries? What if that same worker, preparing for a family team meeting, could ask: What services am I required to offer before considering placement? . . . and then they receive a clear checklist drawn from the actual relevant policies?

1. Child Protective Investigator and Child Protective Investigator Supervisor—Annual Report.

State child welfare agencies can leverage artificial intelligence appropriately to achieve this and many similar objectives. AI can provide a critical resource for overwhelmed frontline workers. The expertise these workers need does not only come from experience—it comes from mastering an impossibly complex web of policies that change regularly, vary by situation, and carry enormous consequences when misunderstood. Child welfare staff have to hold all of this in their heads, simultaneously learning how to react to uncertain and complex situations—all while understanding conditions in a potentially dangerous home, de-escalating a crisis, building rapport, or earning a family's trust.

2. The Fog of Complexity: Synthesizing Information

Mastering relevant policy is only the first layer of the challenge. The situations themselves are complex. A huge portion of "neglect" may arise due to lack of income. A family may need help such as food, housing, or financial support, but not necessarily separation. A caseworker who arrives may see a family struggling but intact, and has to make a call to offer services, close the case, or seek other interventions. After services are provided, that same family may end up in the system again, in similar or perhaps worse circumstances. This impossible calculus faces caseworkers daily.

Scenarios like this are also why documentation became key. Because if things go wrong—if a child is hurt, or even dies—communities will ask: What did the worker know? When did they know it? What did they document?

To address such scenarios, state agencies build a fortress of paperwork. Every visit must be recorded. Every observation catalogued. Every decision justified. For a child in the system for years—bouncing between foster care and kinship placements—that adds up to years of accumulated documentation, including case notes, assessments, interagency meetings, court reports, safety plans, and service referrals. A paper trail meant to protect children and workers alike becomes unmanageable. And buried in those thousands of pages may be information that could save a child's life, such as a pattern of escalating incidents, a detail from a home visit two years ago that suddenly matters now, a note about who the parent turns to when overwhelmed, or a record of what intervention actually worked last time. But how does an overwhelmed caseworker serving 40 families find that crucial detail at 2 a.m. when a crisis hits? They can't, without help.

Now imagine if that same worker could ask, What services has this family received before? ... and then that worker instantly can see a timeline generated from an AI-based resource—not just from their agency, but synthesized across systems? Or, what if a new worker preparing for their first high-risk visit receives an AI-generated brief and asks: What are the three most relevant past incidents? Which intervention approaches have worked with similar families? What are the community resources within two miles of their home? In this scenario, the AI does not make the decision. The worker still knocks on the door, still reads the room, still makes the judgment. But they walk in prepared, supported, equipped—not overwhelmed.

3. Escaping the Documentation Trap

Throughout the 1990s and 2000s, governments digitized many systems. Agencies and industry partners took all that crucial paperwork and moved it into software systems, inadvertently creating a digitization trap. Now workers did not only need to know child development, crisis intervention, and family dynamics. They also needed to understand the complexities of multiple applications. The systems that were supposed to make their jobs easier became another process to follow and another skill to master.

What if, after a difficult home visit, instead of spending three hours writing case notes into a complex application, a worker could speak naturally about what they observed and AI could draft the documentation—capturing details accurately while the worker reviews, refines, and adds irreplaceable human judgment? What if a supervisor, dealing with 50 active cases, could ask: Which families need urgent attention this week? . . . and then they receive information based on appointment dates, missed check-ins, and emerging risk factors—not to make the decision, but to surface what matters most? What if, when a case transfers from one worker to another—as they so often do in high-turnover environments—the new worker could ask: What do I need to know about this family? . . . and then they receive not a mountain of documents to wade through, but a coherent and accurate narrative of what's happened, what's been tried, and what matters most right now.

The Core Paradox

Child welfare offices face a paradox: the need for workers who can read a dangerous situation in seconds, and for workers who can maintain meticulous records. Agencies need practical knowledge and administrative excellence. They need someone who can talk a parent down from a crisis, and someone who can navigate complex policy frameworks flawlessly. They need someone who can build trust with a child in need and someone who can fill out forms perfectly. They are asking for two different people in one body—and often asking the newest, least experienced workers to be both simultaneously.

Until now, this paradox seemed unsolvable. Agencies could not hire twice as many workers. They could not eliminate documentation requirements or simplify the policy landscape—the accountability demands are real, necessary, and not going away. They could not give every new caseworker a seasoned mentor shadowing them every day, answering every question, remembering every detail from every past case. But they can offer help through responsible application of AI.

Roundtable Insights: Government Leaders Seek Sustainable Solutions

A roundtable entitled "The Ethical and Transformative Use of Artificial Intelligence in Child Welfare"—convened by the University of Michigan School of Social Work and the IBM Center for The Business of Government late last year—took place at this moment in the history of the child welfare system. This convening brought together government leaders who have seen this crisis firsthand with experts from academia, nonprofits, and industry.

Leaders responsible for child welfare systems have watched talented workers arrive with passion and leave with burnout. They have witnessed the difficult choices their teams face daily—having to make many complex decisions while facing mountains of data. Despite hiring initiatives, improved training programs, and better supervision ratios, excess turnover persists. The newest workers often still bear heavy burdens.

This roundtable assembled leaders from across the United States (and one internationally), all wrestling with the same fundamental questions: How can government best support these workers in providing services for a high-needs population? How do we give them what they actually need to succeed? Early explorations of AI in child welfare demonstrate that the goal is not to replace human judgment, nor to automate the decision about whether a child is safe, nor to take away the most critical, human parts of this work. The goal is simpler and more profound:

- To remove the part of the job that makes this work unsustainable
- To answer policy questions instantly, so workers can act confidently within legal bounds
- To handle documentation, so workers can focus on the conversation happening in front of them
- To synthesize information across years and systems, so workers can spot patterns that matter
- To prepare workers with context and history, so they walk into every home as informed as possible
- To surface urgent needs, so supervisors can prioritize wisely
- To ensure that when a case transfers, critical knowledge transfers with it

Ultimately, the goal is to free workers from the administrative morass so they can focus on what only humans can do—build relationships, read body language, earn trust, assess danger, keep children safe. Furthermore, AI used as described above can help child welfare agency staff act as the protectors they signed up to be, rather than be prevented from doing so because of the complexity of policy and data entry work that agencies have been forced to create.

This was one of the first roundtables to address AI in the child welfare space, where government leaders come together not just to discuss problems, but to explore solutions that seemed impossible until now. The highlights that follow represent insights from a thoughtful, careful, and ethical exploration of how artificial intelligence might help to reimagine child welfare work. To return to the initial analogy: the goal is not to replace the youngest trees in the forest, but to give them the support they need to grow strong, stay rooted, and actually do the extraordinary work asked of them.

Setting the Stage: Lessons Learned

Participants at this roundtable—including state agencies, large counties and cities, and federal partners—revealed something important: there is tremendous need for this conversation. But there is also tremendous uncertainty. Despite all the interest and potential, there are remarkably few mature applications of AI actually running in child welfare systems today.

This is not for lack of trying. Research on AI in child welfare goes back to the early 2000s: Marshall and English[2] applying artificial neural networks to Washington State's risk assessment model, Schwartz et al[3] using them to predict child welfare and juvenile delinquency outcomes, and then testing them directly against existing tools. But for 20 years, AI in child welfare has mostly been an academic exercise. There have been pilots in New York and scattered attempts across the country, but very few full-scale deployments.

Learning from Controversy

Allegheny County in Pennsylvania remains the most high-profile example, though not without surfacing challenges. The Allegheny Family Screening Tool became a topic in national debate about fairness, transparency, and the ethics of predictive analytics in child welfare. None of the agencies participating in this roundtable expressed any interest in replicating that model. In fact, it became clear that those applying artificial intelligence today are deliberately steering away from allowing AI to make decisions about children and families, learning from its controversies and charting a more cautious, ethically driven path forward.

Indeed, agencies want to learn, and came with open minds and hard questions. Many of them asked about how to adapt use cases for their own states, and how AI might actually help in their specific contexts.

AI Did Not Start with Neural Networks

A historical precedent emerged from the discussions: AI in child welfare did not start with neural networks or large language models. The earliest forms of AI in this field were based on data from paper forms. Structured decision-making tools came into wide use in the late 1980s and 1990's, and are still used today. These tools essentially leverage rudimentary AI: variables, scores, and checkboxes that add up to tell workers if a family is high, medium, or low risk. Or sometimes they simply are checklists that help structure a worker's thinking—a way of representing what an expert, or an all-knowing human, might think about a family, translated into a tool to help that caseworker make an impossible decision.

Understanding this history helped participants see that the question is not whether to use AI. States already do. The question is how to do it better, more ethically, and more effectively. Steven Hintze, Chief Data and Product Officer, Information Technology Data and Product, Arizona Department of Child Safety, put it: "It's like being afraid of the internet. It's not a matter of if, it's a matter of when."

2. Marshall, D.B., and D.J. English. (2000). Neural network modeling of risk assessment in child protective services.
3. Schwartz, et al. (2004). Computational intelligence techniques for risk assessment and decision support.

Hierarchy of AI Use Cases

A clear hierarchy of use cases emerged from the low-hanging fruit to tackle first, and harder problems to approach later.

- **The easiest wins:** Helping new workers understand policies and procedures. If a worker needs to find the right form or understand their agency's policies, AI can help with that. This seemed universally accepted by the group.
- **Next level up—summarization:** Helping workers sift through enormous case files to find what they need, using tools for smart search through documentation. When a child has been in the system for years, the paper trail is massive. AI can help workers find the needle without reading the entire haystack.
- **Administrative hurdles:** Tasks like redacting sensitive information from files, work that currently requires human hours but not human judgment. AI can handle this.
- **Training:** How to prepare these new workers before they are out in the field making real decisions about real families? AI-powered simulations and training scenarios came up repeatedly as promising territory.
- **And then the hardest problem:** Using AI to identify the children at highest risk of serious injury or death. Finding those needles in the haystack. No one wanted to abandon this goal, which is too important. But there was clear consensus that government is not ready yet. Not until agencies have mastered the basics, and implemented the low-hanging fruit.

The group focused on an important distinction buried inside this hierarchy: the difference between what AI should do, and how that gets done. Identifying the right use cases—policy search, summarization, redaction, training—is one decision. Choosing the technical infrastructure to support those use cases is a separate decision entirely. Should the AI run on a commercial platform, accessed over the internet? Or should it run locally, inside the agency's own secure environment, on models the agency controls? These questions carry very different implications for data security, cost, reliability, and long-term sustainability. Too often, conversations jump straight from "AI could help with this" to a specific product or vendor. The roundtable addressed a recognition that agencies need to be deliberate about both questions—first, what problem to solve, and second, how to solve it safely.

Documentation: High-Value Use Case with Critical Caveats

One area that generated particular excitement was documentation. Imagine a worker speaking their case notes after a visit into an application, and AI then drafting the notes. This could save hours of administrative time, freeing workers to spend more time with families and less time behind a screen. But—and this is critical—AI-generated documentation still requires human review. The worker remains the author, and has to verify accuracy, catch hallucinations, and ensure nothing important is missing or wrong. This point came up repeatedly: even the more straightforward uses of AI—finding policies, searching files—still have room for error. Agencies furthest ahead in using AI made this abundantly clear.

Learning from the Leaders: Arizona and Washington, D.C.

Child welfare systems in Arizona and D.C. were advanced in using AI. But these states emphasized that they are still learning and refining. Their agencies just started by tackling areas with the highest administrative burden first—quick wins, places where mistakes do not have high-stakes consequences for children and families.

The need for transparency emerged as a key success factor. The presenters from D.C. and Arizona shared openly with peers, acknowledging challenges and inviting questions. And in that discussion, something became visible: there is no real platform for sharing lessons about using AI to improve child welfare delivery.

A New Approach: Conversation Among Peers

The National Institute of Standards and Technology (NIST) has convened groups around ethical AI. But there's no equivalent forum for addressing ethical AI and child welfare specifically. And this roundtable revealed a need for convenings like this. Agencies operate in silos because there's no infrastructure for continuous learning and collaboration. More importantly, when agencies do come together to discuss AI, it's often at conferences with many sales presentations—companies offering solutions, rather than practitioners sharing effective practices and lessons learned.

At this roundtable, the IBM Center for The Business of Government and the Michigan School of Social Work brought together a different focus. Numerous participants indicated that this was the first time government agencies at vastly different stages of AI adoption—from those just beginning to explore possibilities to those with systems already deployed—came together as peers. The agencies furthest ahead were providing models, and those just starting were asking hard questions that shaped the conversation.

Better communication emerged as a major theme. So did a genuine interest across agencies to learn how AI can make this work more effectively for social workers—to give them more time in the field with families, and less time wrestling with software and paperwork. What emerged was relief at finally having a space where leaders could learn from each other in a non-sales context.

Stakeholder Exclusion: The Ethical Gap

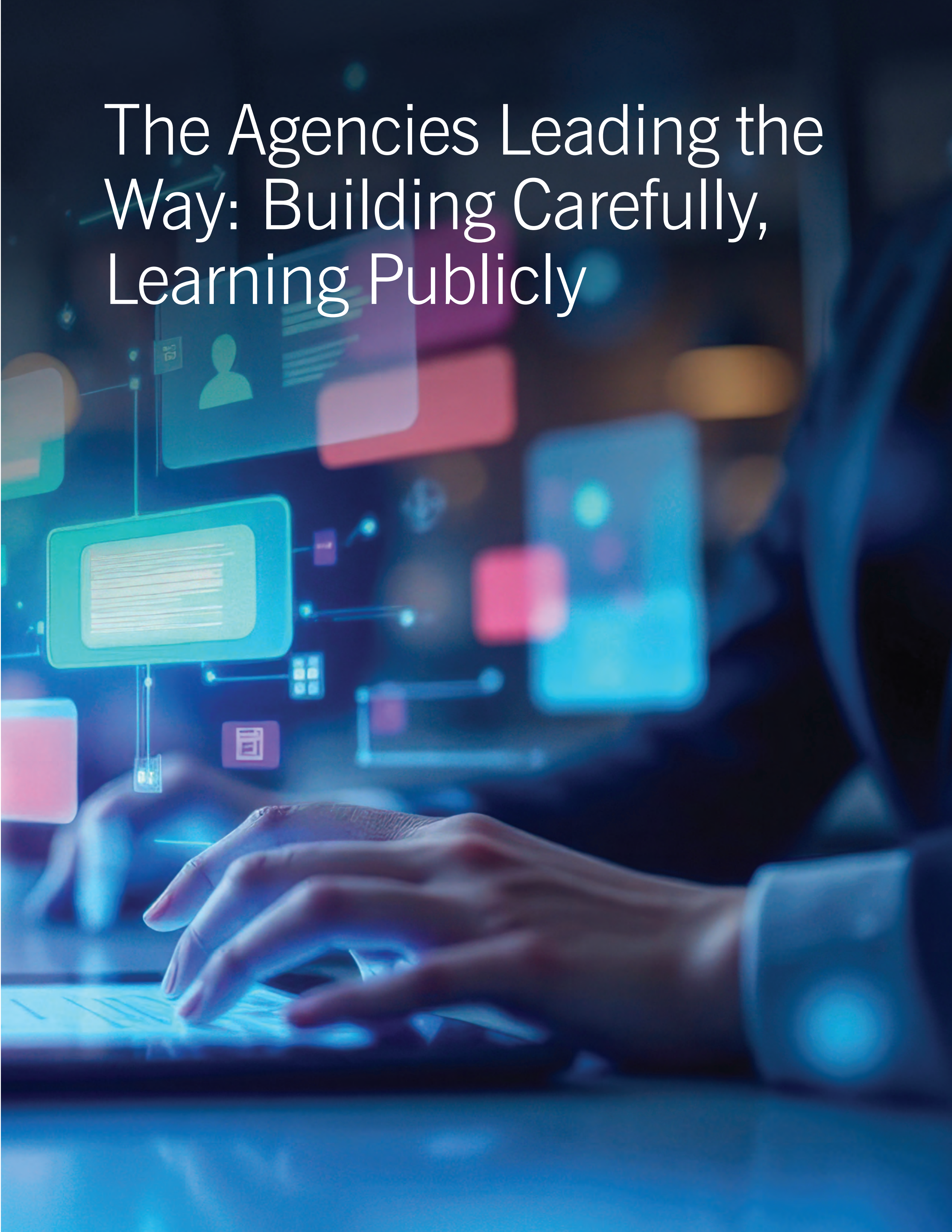
As described by Dr. Brian Perron, professor and co-director at the Child & Adolescent Data Lab, University of Michigan School of Social Work, agencies need to move away from "fear of missing out" with AI. This is not about jumping on a trend, but rather how to apply AI safely and ethically. And ethics came up constantly, especially in the breakout sessions. Agencies are clearly trying to get this right, rather than to cut corners.

One more thing became clear: the academic research on AI that began in the early 2000s has not been brought to bear on current deployment efforts. And when AI pilots are built—sometimes transparently like Arizona and D.C., sometimes behind the scenes—there are usually too few people in the room to scale. Key stakeholders can help to inform perspectives on how best to build effective AI tools, including academic experts as well evaluators, the public, and frontline social workers.

The Path Forward

The conversations reflected in this report represent collective thinking on how to proceed—from government leaders, academics, practitioners, and technologists wrestling with the future. Not by replacing human judgment, but by removing the barriers that keep workers from doing the work that only humans can do. The forest metaphor holds: saplings are unable to bear impossible weight. But they can have better soil, stronger roots, and the support structure needed to grow.

The Agencies Leading the Way: Building Carefully, Learning Publicly



A common theme ran through the day's discussions: the agencies furthest ahead with AI in child welfare are moving carefully, transparently, and feasibly. They are not chasing hype, and they are not replicating the mistakes of the past.

A Different Kind of Leadership

Perhaps the most striking realization from the roundtable was how intentionally agencies like in Arizona and Washington, D.C. have chosen to distinguish their AI work from earlier, more controversial efforts. When Allegheny County in Pennsylvania introduced its predictive screening model several years ago—developed in part by Emily Putnam-Hornstein and Rhema Vaithianathan from New Zealand—it became one of the most high-profile experiments in applying AI to child welfare.

The Allegheny Family Screening Tool used predictive analytics to assess the likelihood of future maltreatment and recommend whether to investigate a case. But critics raised concerns about fairness, transparency, and bias—especially in how predictive systems could reinforce racial disparities and potentially misuse sensitive disability information embedded in historical data. Those conversations still echo across the field. And no one in attendance sought to replicate that model.

The agencies experimenting today are deliberately steering away from starting with AI for predictive decision-making. They are focusing instead on removing administrative burdens, freeing time for workers, and building internal expertise before attempting anything that approaches child-safety decision support. As Steven Hintze from the State of Arizona put it plainly: "We tell our social workers that AI will not replace your job. It will replace parts of the job that involve more routine administrative tasks."

Arizona: The Policy Bot and the Human in the Loop

The clearest example of this philosophy in action came from Steven Hintze and his team at the Arizona Department of Child Safety (DCS). Their work has become a touchstone for how to introduce generative AI into human services in a responsible, measured way.

Arizona's approach began with a simple question: How can AI make it easier for workers to find the information they need to do their jobs? Hintze explained that they wanted to demystify AI and start with "boring" straightforward use cases. Every new case manager must navigate an enormous policy manual—the backbone of agency operations and compliance. Knowing where to find the right procedure or clarification can take valuable time away from families.

The Arizona team designed a policy bot, a smart search tool embedded within their case management system, capable of interpreting natural-language questions and returning precise policy guidance. The bot does not summarize policy or paraphrase—it surfaces the exact section to answer a query, verbatim, from the agency's official manual. A worker might ask: How do I drug test a parent? . . . and the system will respond by confirming the intent of the question—Do you mean a parent, a child, or an employee?—before retrieving the relevant policy.

It is designed with a human-in-the-loop structure: the worker validates the question, confirms the context, and decides whether to act on the guidance. Hintze described it simply: "It's not going to make a safety decision. It's not going to decide permanency." It will help a worker do their job with less friction. The underlying model runs inside Arizona's own secure cloud computing platform, meaning no data leaves their environment.

The state has built in disclaimers against entering personally identifiable information, and has embedded controls to prevent the model from "learning" on sensitive data. The design philosophy was to build internal capability, not dependency on an external source. Greater capability drives greater effectiveness. Put another way: "You can do cloud poorly or locally poorly," Hintze said, "or you can do both well."

Lessons from the Arizona Team

The Arizona demonstration generated high engagement—not because it was flashy, but because it was practical. Participants were struck by how much intentional design had gone into applying AI responsibly. The Arizona agency began by defining what success would look like:

- Not replacing people, but replacing the parts of the job that drain people
- Not automating judgment, but automating access to information
- Not experimenting in the dark, but documenting everything—what works, what doesn't, and why

One attendee asked whether Arizona's ethical guidelines were made available for peer review, as a model for states drafting ethical guidelines and policies for AI. Hintze explained that Arizona established a governance council reporting directly to the agency director, with representatives from across divisions. This group meets monthly to review new ideas and reports to the director–to audit unstructured data use, and ensure every pilot has clear ethical oversight. Arizona also built an internal peer-review process modeled loosely on IBM's AI Ethics Board, creating a homegrown system of accountability that can evolve with experience.

Other lessons learned were candid and valuable. When the agency upgraded to a new Gen AI model, accuracy initially dropped. The team had to conduct extensive user acceptance testing (UAT) to recalibrate prompts and restore trust in the tool. They discovered that even small updates could alter tone, mannerisms, and confidence levels—proof that these systems need continuous evaluation. They learned that tracking metrics mattered: which questions workers were asking, how often the bot was right, and when it failed. The team tracked "rage clicks," a small but telling metric showing when workers clicked repeatedly out of frustration—an indicator of where to improve. By building feedback loops into the pilot itself, Arizona's DCS didn't just test a product; it tested an approach to cultural adoption. It built trust by treating caseworkers as co-designers.

An IT administrative leader shared that his state was currently "crawling and learning to walk" when it comes to AI, and appreciated the opportunity to learn from the group how this could be applied to child welfare successfully.

Washington, D.C.: From Documentation to Design

If Arizona's story was about precision and governance, D.C.'s was about breadth and creativity. Samira Alikadeiyeva, from D.C.'s Child and Family Services Agency, described a similar philosophy—taking away the time-consuming and repetitive administrative components of the job, starting small, focusing on the human interface with AI, and starting with what workers actually need.

Working as a child protective investigator in D.C. is a high-pressure job in a high poverty and high-needs city—and getting workers up to speed is a difficult task. The team working on D.C.'s applications of AI were fully aware of how much time social workers spent "bogged down," looking through case files, "trying to find information in years-long cases."

Washington, D.C. is deploying a new, agency-wide child welfare information system. The first step in this process was digitization: turning paper into screens. Once that foundation was in place, the team began to explore where AI could make the biggest impact. Their first use case mirrored Arizona's—policy search and guidance. With a new system and hundreds of new workers coming on board, it was critical to make policy access frictionless—as well as helping workers learn a new child welfare information system. The developers of the first use cases simply wanted to give workers a "leg up" in addition to the training they received. In theory, the AI application would support the training, and serve as an expert to advise workers. Initial adoption of the agency's AI-powered 'Case Agent' was gradual, as staff adjusted to a new system and associated changes in workflow. The lesson was clear: timing matters. AI cannot succeed if underlying workflows are not yet stable.

The agency narrowed down from a list of 30 to just 4 use cases, with others in the pipeline. Some are more advanced than others.

One key use case for the agency is Case Agent, an AI-generated agent that quickly gives workers answers on a particular case. This is critical given the volume of information that can be collected for a single child in care. Alleviating this administrative burden placed on workers gives them time back to think about how to meet the needs of children in care. While not live yet, this use case could make a caseworker's job more manageable.

Contact notes reflected another critical use case for the D.C. agency. Some case workers, while they are in and out of their cars for meetings, prefer to write their contact notes on paper. AI-enabled contact notes allow busy workers to take a quick snapshot of their written notes, and have AI generate typed summaries required for the child welfare information system—cutting tremendous amounts of time and energy typing these notes. Double data entry is eliminated. The agency is now testing this with a group of social workers, and it has generated a great deal of interest.

Workers can also dictate notes into their agency-issued device, record them, and AI will generate typed and cleaned up, professionally formatted contact notes—ready to be validated and edited by the worker, before being ingested into the child welfare information system rapidly and seamlessly.

Again, this approach can save time. The agency realized that contact notes represent the lion's share of the documentation needed in the high turnover job of child protective investigator facing "huge caseloads" in Washington, D.C. Adding an investigation summary capability to rapidly document and transcribe contact notes could be a game changer, and that could be replicated across the United States.

The team also developed an AI-driven service request assistant that scans court orders and automatically drafts court-ordered service requests, helping workers keep up with the myriad of required documentation and tasks needed. Given the complex needs of families in the city, time is of the essence and AI can help.

Other pilots focused on matching children and families to services—from therapy to housing—using data-driven recommendations, while keeping human decision-making intact.

Marina Havan, from the Washington, D.C. team, pointed out that the D.C. child welfare system is excited about implementing multiple use cases developed by social workers, and focused on a return on investment. Havan advised that this needs to be a cost vs. benefits analysis, and sometimes it is difficult to ask for money for AI projects based solely on the metric of saving time.

It was also highlighted that implementing AI use cases is not simply about identifying the best use cases. This also involves incorporating change management as part of a holistic approach to AI adoption.

A Shared Philosophy: Low-Risk, High-Reward

Both Arizona and D.C. are modeling what responsible innovation looks like in public human services. They are not implementing AI because of a fear of missing out. They are not holding AI like a hammer and looking for a nail. They are solving for time, clarity, burden, and relief. Overall, participants shared a collective understanding that automation should relieve administrative pain, not create fear.

Throughout the discussions, agency leaders and academics echoed similar sentiments. Dr. Brian Perron, with The Child and Adolescent Data Lab at the University of Michigan (the DataLab), outlined a disciplined approach to adopting AI responsibly: document inefficiencies, establish baselines, evaluate performance against human benchmarks, and calculate error before scaling anything.

Dr. Perron also advised the attendees to focus on uncovering rich context from unstructured data using AI, and described it as a low risk, high reward component that could complement a number of different use cases. In Michigan, the DataLab has been supporting the state with AI tools for gaining insights into their data.

Others highlighted the importance of data privacy. Cloud computing models are not private by default. Local deployment or tightly controlled cloud environments, like those that Arizona has done, offer the control necessary to protect sensitive child welfare information.

The Data Lab, for example, has taken this principle further, committing to using exclusively local, open-source models for all research involving confidential child welfare data. Local deployment means no confidential information ever leaves the secure computing environment—not a redacted version, not a summary, not a query. It also means the research team controls which model version runs at all times. There are no surprise updates, no shifts in how the model interprets a question, no risk that a validated workflow stops working because a provider changed something upstream. The model tested is the model that runs.

Some believe that local models cannot match the performance of larger, commercial systems. Recent evidence suggests otherwise. Since the roundtable convened, the Data Lab published a systematic evaluation of models ranging from 0.6 billion to 32 billion parameters on four child welfare classification tasks: identifying substance-related problems, domestic violence, firearms, and opioids in case narratives (Qi et al., 2026).[4] A 4-billion-parameter model—small enough to run on consumer-grade hardware costing approximately $2,000 to $3,000—achieved almost perfect agreement with expert human coders, with Cohen's kappa values between 0.93 and 0.96 on three of four benchmarks. It outperformed models up to eight times its size. The performance gap that once justified reliance on larger, cloud-based systems has narrowed to the point where it may no longer exist for many practical child welfare applications.

Building Toward Collaboration

A rich exchange among participants followed these demonstrations. One state recognized that workers are asking for AI tools to make their lives easier. Agency leaders seemed to agree that the correct approach was "learning to walk before running with AI"—starting with small, visible wins that build confidence. Another State's team emphasized that the goal was not just to save time, but to give workers more presence with families.

Academics in attendance urged agencies to bring researchers and evaluators into the room early. "We don't want isolated pilots," one professor said. "We want learning systems." Others emphasized human-centered design—the slow but necessary process of co-developing tools with the people who will use them.

An important theme emerged from the group's focus on AI for social workers: what if AI could be placed safely and thoughtfully in the hands of a 15-year-old in foster care to help them navigate the extraordinary challenges they face? The IBM Center for The Business of Government offered an example from the Department of Veterans Affairs, highlighting AI-supported self-help tools for veterans. Such a tool might be adaptable to support children in the foster care system are a notable exception.

Another topic of general agreement involved the issue of how to develop AI within an agency. While it might be easier to navigate the evolution of AI with a few data scientists and/or researchers, subject matter experts and frontline social workers and their supervisors need to be included early and thoughtfully. This means at the concept phase, through training, testing, bias evaluation, and user acceptance. The group talked at length about the importance of AI-enabled case summaries, and social workers, subject matter experts, and social science researchers can all play a role in case summary validation—to make sure AI is not hallucinating and providing overwhelmed frontline workers with bad information. In other words, it is not as simple as going off in isolation and returning with a model—it takes a multi-disciplinary team to do this correctly.

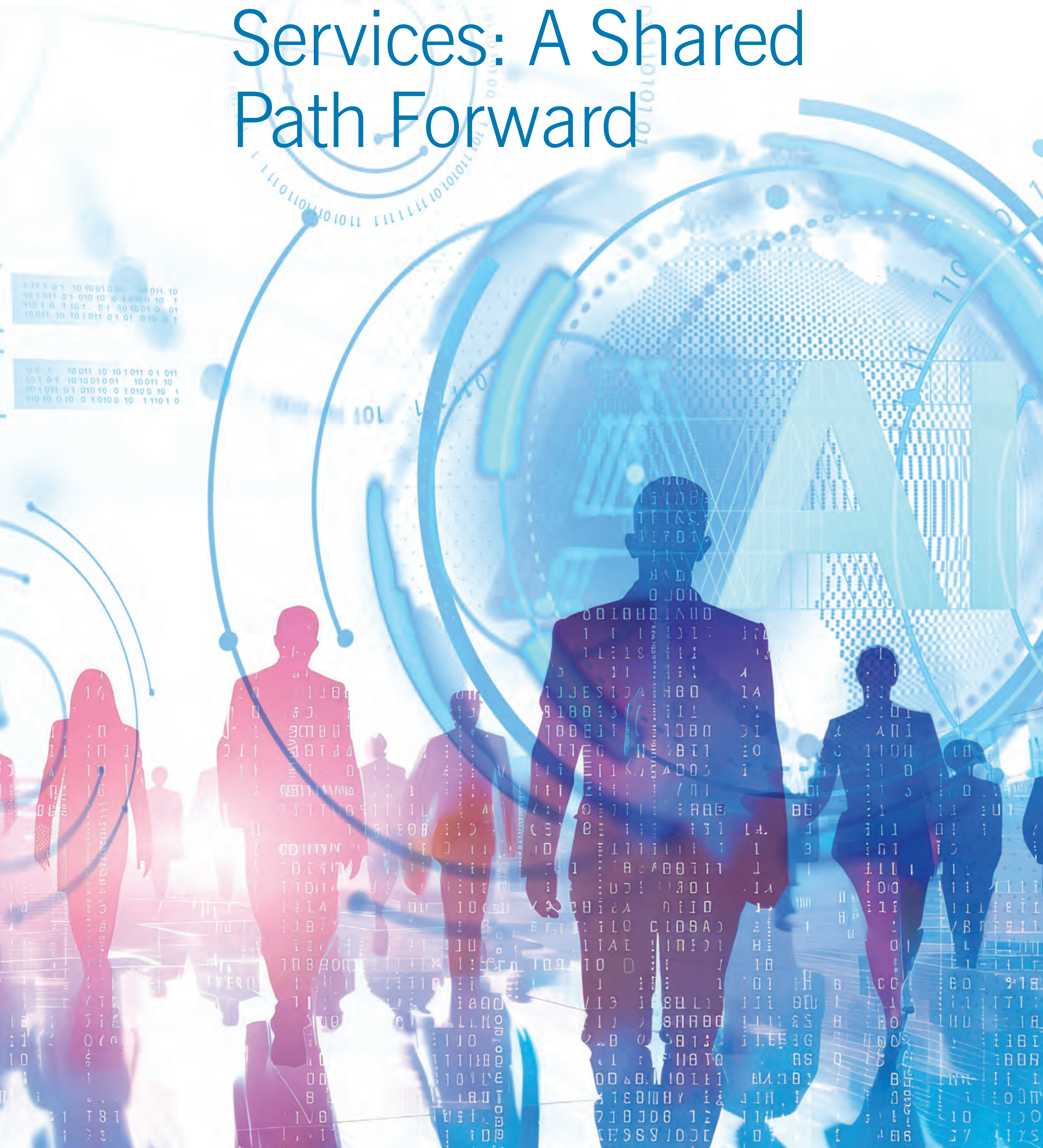
4. Qi, Z., B. E. Perron, B. G. Victor, D. Stoll, and J. P. Ryan. (2026). Small models achieve large language model performance: Evaluating reasoning-enabled AI for secure child welfare research. *Journal of Evidence-Based Social Work.*

What Leadership Looks Like in This Moment

The roundtable discussion demonstrated how to take the first steps responsibly. Leading states like Arizona and Washington, D.C. show that low-hanging fruit—policy search, document summarization, redaction, training simulations—can have profound impact without crossing ethical boundaries. Yet there is no NIST-equivalent for child welfare, no national hub for sharing lessons, governance models, or open frameworks. Agencies are learning in parallel instead of in concert.

The roundtable made clear that state agencies can benefit from shared infrastructure for shared learning—a network where states and localities can compare approaches, publish results, and learn from one another, and where progress is possible without perfection. When agencies move deliberately, when they involve their workforce, when they prioritize the right problems, AI can become a partner instead of a challenge. It can free caseworkers to spend less time wrestling with drop-down menus and more time sitting with families—listening, observing, deciding, and protecting.

The Future of Human Services: A Shared Path Forward



The discussion concluded with open conversation about where agencies are in their AI journey, what they are learning, what is working, and what remains hard. Agencies were not considering AI as an abstract idea anymore. They were talking about where AI fits, what it can safely take off workers' plates, and how to make sure the technology serves the mission. The following themes capture that discussion—key takeaways about where AI is already helping, where the limits are, and how to move forward in support of the people doing the work.

From Curiosity to Capability

Across the country, the question about AI has shifted. It's no longer "should we use it?" but "how do we use it responsibly?" Agencies are beginning to take inventory of inefficiencies, document performance baselines, and test AI tools against real human benchmarks. The goal is not to chase new technology, but to understand where it can help. This shifts the focus from experimentation to readiness—knowing how to measure results, account for risk, and define the parts of the job that will always require human judgment. Early discussions about AI in child welfare often centered on whether it was appropriate at all. Now the conversation has evolved to *how*—how to implement it safely, how to measure its impact, and how to ensure it serves workers rather than burdening them.

Relief, Not Replacement

One theme came up repeatedly: AI should make the job less bureaucratic, not smaller. AI can relieve workers of the repetitive tasks that keep them from the work that actually matters. That means automating data entry, compliance forms, and routine documentation, so caseworkers can spend more time in the field and less time behind a screen. The work is not about replacing people, but instead about giving them back time to focus on service. This philosophy echoes the forest metaphor from the introduction. AI is not about removing the saplings—it's about giving them better soil.

Humans in the Loop

Every example shared at the roundtable reinforced the same point: humans must stay in control. No AI system should ever make a safety or permanency decision. Tools can summarize, search, or draft, but only people can interpret, empathize, and decide. Several states have begun to formalize this through ethics councils and internal governance boards that review each project before deployment. These structures give agencies a way to test, monitor, and audit AI systems instead of relying on outside vendors. Trust is not a feature of the software; it comes from oversight and accountability. Arizona's governance council, meeting monthly with cross-divisional representation, became a model that other agencies can follow—having a regular, structured process for asking hard questions before problems arise.

Building Trust Through Transparency

Trust and transparency go hand in hand. Workers need to understand where an answer comes from, not just see the output. Arizona's internal policy search tool became a key example. Workers type questions, confirm what they mean, and receive the exact policy language rather than a paraphrase. Accuracy is tracked, feedback is logged, and all stays within a secure state environment. That kind of clarity builds confidence—it shows workers that the tool is there to help, not to guess. The contrast with earlier AI efforts was stark. Where previous systems often felt like black boxes, these new tools emphasize explainability. Workers can trace the path from question to answer, verify the source, and maintain their professional judgment throughout.

Privacy, Data, and Local Models

Privacy came up in nearly every discussion. Most agencies prefer local models that stay inside their firewalls, instead of relying on public cloud services. The basic rule is simple: keep the data where it lives, and share the algorithm. This approach allows agencies to use generative AI while staying compliant with HIPAA and state privacy laws. It also avoids recurring API costs and limits the risk of data exposure.

Even so, no one described the tools as perfect. Current redaction models are about 80 percent accurate in removing private or health-related data, so human review remains essential to balance automation and verification.

The technical architecture matters because the stakes are so high. These are not consumer applications where errors mean inconvenience. These systems handle highly sensitive information about vulnerable families.

The choice between frontier models and local models warrants attention. Frontier models in wide use today (e.g., Chat GPT or Gemini) are powerful, easy to access, and require no specialized technical infrastructure. For tasks that do not involve confidential data, such as drafting training materials or summarizing publicly available research, they can be practical and effective. But when confidential case records are involved, the calculus changes. Every query sent to a frontier model transmits data to an external server. Even with contractual safeguards, the organization can cede physical control of that information.

Participants also discussed the question of model stability. Frontier model providers update their systems routinely, sometimes without notice. An update can change how a model responds to the same question, altering classification accuracy or output formatting in ways that affect previously validated workflows. Agencies that depend on a frontier model for consistent, repeatable results must therefore continuously revalidate their tools after every update—an often impractical maintenance burden. Local models, by contrast, do not change unless the organization deliberately changes them. The version that was tested and validated is the version that runs, today and tomorrow.

Cost was identified as a third factor. Frontier models accessed through APIs charge per unit of text processed. For a single query, the cost is trivial. For tens of thousands of case records, it adds up fast. Local models, once deployed on the organization's own hardware, carry no per-record processing cost. For agencies conducting large-scale analysis of administrative records, this difference is not marginal—it is the difference between a sustainable program and one that requires ongoing budget justification for every run.

Real Work, Real Progress

The examples shared were practical and focused on solving specific problems.

- Arizona built a policy search bot that makes it easier for workers to find accurate information quickly and securely.
- Washington, D.C. focused on AI-assisted documentation, allowing caseworkers to dictate or summarize notes and service requests while keeping ownership of the final text.
- Virginia is testing AI for audit support, translation, and accessibility, with an emphasis on improving consistency and reducing administrative delays.

Each effort started small and aimed at the same goal: remove friction first. These were not moonshot projects. They were targeted interventions addressing specific pain points that workers identified themselves.

Challenges Still Ahead

Despite progress, the challenges were consistent across agencies. Trust takes time. Usability remains an issue; tools must fit into existing case systems. Workers still worry about being replaced. Pilots are fragmented and rarely coordinated across states. Ethical validation processes are inconsistent. Feedback cycles are slow. Technology that is not shaped by the people who use it can contribute to another layer of bureaucracy. As one person said, "We don't need more solutions sitting on a shelf." Attendees did not gloss over difficulties, but rather acknowledged what wasn't working yet and why.

From Fear to Fluency

Success will depend on how comfortable the workforce becomes with these tools. Social workers are already starting to use prompt-based systems, learning to ask questions, refine responses, and verify results. Training and safe testing environments are essential. As workers grow more confident, fear turns into fluency. The cultural shift is as important as the technical one. Workers need time to experiment, make mistakes in low-stakes environments, and build muscle memory with these tools before using them in high-pressure situations.

Connecting the System

As agencies gain experience, the next step is connecting systems across government. AI's potential grows when it links programs like Medicaid, SNAP, and child welfare to serve families more holistically. Virginia's work in this area showed how integration can reduce duplication and help families access services faster. This can increase both efficiency and fairness—making sure families get help before problems escalate. This represents a shift from service within silos to coordination across them. The vision is not just a more efficient child welfare system, but a more integrated safety net that can identify and respond to family needs earlier.

Conclusion and Recommendations

Roundtable attendees agreed on a clear message: everyone is learning this together. No state has a perfect model, but each agency is contributing a piece of the puzzle. Every pilot, even the imperfect ones, adds to the field's understanding. The next step is to connect those lessons—to build a shared community of practice so progress in one place speeds up progress everywhere.

The Need for Federal Support and Shared Infrastructure

Federal mechanisms already exist for launching innovation projects, and they could be adapted to support AI readiness and collaboration in human services. State agency leaders responded positively to the idea. They saw how federal support could fund shared infrastructure, connect state and local pilots, and create a structured way for agencies to learn from each other instead of working in silos. The goal would not be to centralize control, but to create a common platform for communication, technical assistance, and evaluation—a bridge between experimentation and coordinated learning. This kind of structure could turn today's scattered pilots into a connected network for progress. It would provide what the roundtable revealed was desperately needed: a way for agencies to learn together, rather than repeat mistakes in isolation.

Principles To Build On for Using AI in Child Welfare

The discussion pointed to ten principles that agencies at every stage—from those just beginning to explore AI to those with systems already deployed—could embrace.

- Relief, not replacement
- Humans in the loop always
- Start with low-risk, high-reward
- Workers as co-designers
- Transparency and explainability required
- Local control of sensitive data
- Multi-stakeholder development
- Continuous evaluation and learning
- Bias monitoring throughout
- Families and youth as stakeholders

These principles are not prescriptions imposed from above. They emerged from observations about what's working, distilled from candid conversations among practitioners and partners. They represent a foundation that agencies can build on, adopt formally, and refine through their own experience. They mark a clear path forward—one that learns from past controversies, while embracing AI's potential to make a demanding job more sustainable.

The Path Forward

Throughout the day, a common refrain emerged: appreciation for finally having a space to learn together. The saplings in this forest still bear heavy fruit. But they need not do so alone. The discussion showed how to build tools that support rather than burden workers. States are leading, following, learning, and adapting. The question is no longer whether AI belongs in child welfare, but how quickly agencies can share what works so that every worker, in every state, can have the support they need from day one. The forest is still young. But it's growing stronger—together.

References

Child Protective Investigator and Child Protective Investigator Supervisor - Annual Report. (2024, October 1). Department of Children and Families Office of Child & Family Well-Being.

Marshall, D. B., and D.J. English. (2000). Neural network modeling of risk assessment in child protective services. *Psychological Methods*, *5*(1), 102–124.

Schwartz, D. R., A. B. Kaufman, and I. M. Schwartz. (2004). Computational intelligence techniques for risk assessment and decision support. *Children and Youth Services Review*, *26*(11), 1081–1095.

Qi, Z., B. E. Perron, B. G. Victor, D. Stoll, and J.P. Ryan, (2026). Small models achieve large language model performance: Evaluating reasoning-enabled AI for secure child welfare research. *Journal of Evidence-Based Social Work*. https://doi.org/10.1080/26408066.2026.2616711.

Vaithianathan, R., N. Jiang, T. Maloney, P. Nand, and E. Putnam-Hornstein. (2017). Developing predictive risk models to support child maltreatment hotline screening decisions: Allegheny County methodology and implementation. Centre for Social Data Analytics, Auckland University of Technology.

Vaithianathan, R., E. Kulick, E. Putnam-Hornstein, and D. Benavides-Prado. (2019). Allegheny Family Screening Tool: Methodology, Version 2. Centre for Social Data Analytics, Auckland University of Technology.

Associated Press. (2022, April 29). How an algorithm that screens for child neglect could harden racial disparities. PBS NewsHour.

California Department of Social Services. (n.d.). Structured Decision Making Project (Initiated 1998). Retrieved from https://www.cdss.ca.gov/inforesources/child-welfare-protection/structured-decision-making.

De-Arteaga, M., S. Feuerriegel, and A. Praetzel. (2024). The Allegheny Family Screening Tool's overestimation of utility and risk. *Logic Magazine*. https://logicmag.io/policy/allegheny-family-screening-tools-overestimation-of-utility-and-risk/.

Evident Change. (n.d.). SDM® model in child protection. Retrieved from https://evidentchange.org/assessment/structured-decision-making/child-welfare/.

Samant, A., N. Shemtov, K. Xu, S. Beiers, M. Gerchick, T. Jegede, T. Shah, A. Gutierrez, and A. Horowitz. (2023). The devil is in the details: Interrogating values embedded in the Allegheny Family Screening Tool. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1292–1310). ACM.

Note on AI Use

This report was authored by David R. Schwartz. All content, themes, findings, analysis, and conclusions are my original work based on the roundtable discussions and my synthesis of participant contributions. Generative AI tools were used only for limited editorial feedback and proofreading assistance.

About the Author

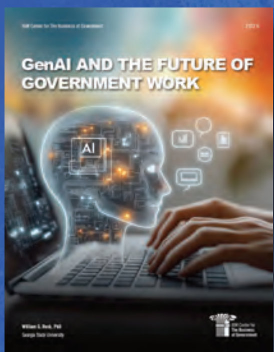
David R. Schwartz, MSW
AI Policy Practitioner & Social Work Researcher
Naples, Florida, United States

LI: linkedin.com/in/david-schwartz-msw-22467022
E: david.ross.schwartz@gmail.com

David R. Schwartz is a social work researcher and AI policy practitioner whose work at the intersection of technology and human services spans more than two decades. A former front-line child welfare worker, he led one of the first formal evaluations of artificial intelligence in child welfare and was among the first researchers to apply artificial neural networks to predictive modeling in child welfare and juvenile justice—foundational work begun in 1999 that helped establish early principles for computational intelligence in social services.

His career has encompassed large-scale technology implementation for child welfare agencies, strategic data infrastructure development, and sustained policy guidance on the ethical deployment of AI across vulnerable population services. He holds a Master of Social Work degree in macro practice from the University of Pennsylvania School of Social Work. Mr. Schwartz contributed to this roundtable bringing rare practitioner depth—grounded in both direct service experience and decades of research—to questions of responsible innovation in public child welfare systems.

Recent Reports from the IBM Center for The Business of Government



GenAI and the Future of Government Work

by William G. Resh

Outcome-Based Contracting in U.S. Government

by Daniel Finkenstadt and Tim Cummins

Responsible AI for Public Evaluation

by Daniel Fonner

Practical Cyber Solutions for Managing Government Supply Chains

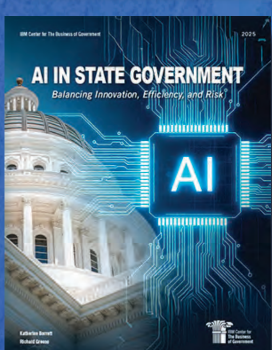
by Robert Handfield

Five Pillars of Effective Government

by Dan Chenok

Government's Digital DNA: Identity and Access Management for Public Sector Security

by Andrew Whitford



AI in State Government

by Katherine Barrett and Richard Greene

Building community-based resilience

by Authors of Case Study

For a full listing of our reports, visit businessofgovernment.org/reports

About the IBM Center for The Business of Government

Through research stipends and events, the IBM Center for The Business of Government stimulates research and facilitates discussion of new approaches to improving the effectiveness of government at the federal, state, local, and international levels.

About IBM Consulting

With consultants and professional staff in more than 160 countries globally, IBM Consulting is the world's largest consulting services organization. IBM Consulting provides clients with business process and industry expertise, a deep understanding of technology solutions that address specific industry issues, and the ability to design, build, and run those solutions in a way that delivers bottom-line value. To learn more visit ibm.com.

For more information:
Daniel J. Chenok
Executive Director
IBM Center for The Business of Government

600 14th Street NW
Second Floor
Washington, D.C. 20005
(202) 551-9342

website: www.businessofgovernment.org
e-mail: businessofgovernment@us.ibm.com

IBM Center for
The Business
of Government