# **Del hype a la política pública**

# Cómo gobernar la IA en el sector público

# Lecture Outline (ESAP Manizales)

## Segment 1: **Introduction – AI’s Global Rise and Why It Matters** (15 minutes)

**Focus:** Setting the stage with AI’s rapid integration into everyday life and the big questions it raises for society and governance. Introduce the key themes and why public administrators and students should care.  
**Main Ideas:**  
- **AI’s ubiquity:** Artificial intelligence has moved from research labs into daily life at remarkable speed, transforming industries and public services[[1]](https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md#L20-L28). From smartphone assistants to smart city systems, AI is all around us.  
- **Opportunities vs. concerns:** Alongside exciting innovations (e.g. smarter healthcare, efficient transport) come **profound questions** about oversight, ethics, and societal impact[[1]](https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md#L20-L28). The audience’s dual perspective (students and public sector) positions them to both use AI and regulate it.  
- **Global push for responsible AI:** Around the world, governments and organizations are grappling with how to ensure AI serves the public good while minimizing harms[[1]](https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md#L20-L28). This talk will explore how different regions (Canada, EU, Brazil, etc.) and Colombia itself are approaching this challenge.  
- **Overview of themes:** Briefly preview the lecture’s five core topics – AI Policy & Regulation, AI in Healthcare, Autonomous Vehicles, Algorithmic Fairness, and AI & Society – which together highlight the interplay between technology and public policy.  
**Recommended Visual Aids:**  
- A world map highlighting countries with national AI strategies or laws (e.g. EU’s AI Act, Canada’s directive, Brazil’s AI bill).  
- A slide with headline examples of AI in daily life (news clips or images: e.g. a phone with a virtual assistant, a self-driving car test, etc.) to illustrate AI’s reach.  
- A quick statistic graphic (for example: “**90%** of leading businesses are investing in AI” or “Billions of AI-powered devices worldwide”) to underscore AI’s growth.  
**Engagement Idea:** Start with a quick audience poll: “**Raise your hand if you’ve interacted with an AI tool (like ChatGPT or Siri) today.**” (Expect many hands, showing AI’s prevalence.) Follow-up with: “How many of you trust AI’s decisions as much as a human’s?” – a question to provoke reflection on trust and set up the discussion to come.

## Segment 2: **AI Policy and Regulation – Balancing Innovation and Protection** (15 minutes)

**Focus:** How governments and institutions around the world are creating policies to govern AI. Emphasize the balancing act between fostering innovation and protecting citizens’ rights and safety.  
**Main Ideas:**  
- **Why regulate AI?** AI can influence critical decisions (jobs, healthcare, security), so oversight is crucial. Unchecked AI might cause harm (bias, privacy invasion, safety risks), but over-regulation could stifle innovation. The key question: *How do we get it “just right”?*  
- **Global approaches:** Compare leading regulatory frameworks:  
- **European Union:** Taking a comprehensive approach with the EU AI Act – the world’s first broad AI law. It uses a **risk-based model**: AI systems are categorized as *“unacceptable risk” (banned), “high-risk” (strict requirements), “limited” (transparency), or “minimal risk”*[[2]](https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md#L32-L40)[[3]](https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md#L40-L43). (E.g. social scoring is banned; medical AI or self-driving cars are “high-risk” with heavy oversight.)  
- **Canada:** Integrating AI governance into existing laws. For example, Canada’s government requires an **Algorithmic Impact Assessment (AIA)** before deploying any AI system in public services[[4]](https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md#L76-L84)[[5]](https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md#L86-L89). This means officials must evaluate an AI’s potential bias, explainability, and privacy impacts *in advance* – a proactive approach to accountability.  
- **Brazil:** Emerging economy perspective – Brazil’s Senate approved a national AI Bill in 2024 focusing on accountability and ethics in AI[[6]](https://www.mattosfilho.com.br/en/unico/framework-artificial-intelligence-senate/#:~:text=On%20December%2010%2C%202024%2C%20the,into%20law%20and%20take%20effect). It proposes a Europe-like risk classification: *“excessive risk”* AI (e.g. coercive or crime-predicting algorithms) would be prohibited, while *“high-risk”* AI (in critical infrastructure, justice, health, autonomous vehicles, etc.) must meet strict requirements like human oversight and algorithmic audits[[7]](https://www.mattosfilho.com.br/en/unico/framework-artificial-intelligence-senate/#:~:text=infrastructure%20,of%20affected%20individuals%20or%20groups). This shows even outside Europe/North America, AI governance is a priority.  
- **Other regions:** The U.S. is currently using sector-specific guidelines (no single AI law yet), and China is mandating security reviews and aligning AI with state interests. These contrasting philosophies (laissez-faire vs. controlled) offer a chance to discuss what might work best for Colombia.  
- **Key regulatory issues:** Highlight recurring themes in AI policies: transparency requirements (e.g. informing people when AI is used), safety standards, data privacy (AI driven by big data must respect privacy laws), and accountability for harms. Many laws insist on a *“human in the loop”* for important decisions[[8]](https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md#L134-L142) to ensure human control.  
- **Real-world examples:** Mention cases like the EU’s draft AI Act influencing other nations, or how New York City now requires bias audits for AI hiring tools (one of the first local laws on algorithmic fairness[[9]](https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md#L100-L103)). These illustrate governments reacting to specific issues with AI.  
- **Relevance to Colombia:** Note that Colombia has a national AI strategy and is part of global forums on AI. Ask: *How might Colombia’s context (e.g. institutional capacity, social priorities) shape its AI regulations?*  
**Recommended Visual Aids:**  
- **Chart/Map:** A comparative chart of AI regulatory approaches (EU vs US vs China vs Latin America) – e.g., a table or infographic with columns for each region and rows for key aspects (strategy vs. specific laws, strictness level, examples of rules).  
- **Timeline:** A timeline of major AI policy milestones (2017 Canada’s AIA, 2021 draft of EU AI Act, 2024 Brazil AI bill, etc.) to show how quickly this policy area is evolving.  
- **Headline screenshot:** e.g. a news headline: “EU Passes Landmark AI Law” or “Brazil Proposes New AI Regulation” – to ground the discussion in current events.  
**Engagement Element:** Pose a scenario to the audience: *“Imagine you’re a policymaker: an AI system in a hospital misdiagnosed several patients. What regulations or actions would you push for to prevent this?”* Allow a couple of audience members to respond. Alternatively, a quick poll: “**Who thinks we need strict AI-specific laws, and who thinks existing laws (like consumer protection) are enough?**” (Show of hands for each side.) This engages the room in the core debate of this segment – how much and what kind of regulation is needed.

## Segment 3: **AI in Healthcare – Transforming Medicine, Saving Lives** (15 minutes)

**Focus:** Exploring how AI is revolutionizing healthcare and public health, while examining the policy and ethical implications. Emphasize both the life-saving potential and the need for oversight to ensure equity and safety in health applications.  
**Main Ideas:**  
- **Revolutionizing care:** AI is being used to **diagnose diseases from medical images, predict outbreaks, personalize treatments, and discover new drugs**. From diagnostics to drug discovery, *no aspect of medicine is immune from being transformed by AI*[[10]](https://www.thinkglobalhealth.org/article/health-latin-america-and-promise-artificial-intelligence#:~:text=,in%20Latin%20America). For example, AI systems can analyze X-rays or MRIs faster (and sometimes as accurately) as specialists, helping doctors catch problems earlier. During COVID-19, AI tools helped track virus spread and even assist in vaccine research.  
- **Benefits in context:** AI could be a game-changer for regions with limited medical resources. The G20 recently noted AI *“holds immense potential to bridge health disparities…personalizing care, enhancing diagnostics, improving efficiency, boosting patient engagement, and expanding telemedicine”*[[11]](https://www.thinkglobalhealth.org/article/health-latin-america-and-promise-artificial-intelligence#:~:text=But%20interest%20in%20health,patient%20engagement%2C%20and%20expanding%20telemedicine). In practical terms, this means an isolated clinic in rural Colombia might use an AI diagnostic tool to identify illnesses in patients where no specialist is on-site – potentially **saving lives by augmenting scarce healthcare staff**. Latin America’s health sector is looking to these innovations to bolster strained systems.  
- **Latin America initiatives:** Awareness of AI’s promise in health is growing across the region[[12]](https://www.thinkglobalhealth.org/article/health-latin-america-and-promise-artificial-intelligence#:~:text=Awareness%20of%20AI%27s%20potential%20for,for%20the%20region%27s%20urban%20health). Health ministries in **Colombia, Brazil, Argentina, Chile** and others have launched initiatives to apply AI in healthcare[[12]](https://www.thinkglobalhealth.org/article/health-latin-america-and-promise-artificial-intelligence#:~:text=Awareness%20of%20AI%27s%20potential%20for,for%20the%20region%27s%20urban%20health) – for instance, national networks to promote telemedicine, or pilot projects using machine learning to predict disease outbreaks. **Brazil** currently leads with an ambitious health AI strategy, investing in R&D; Brazilian scientists even helped demonstrate AI’s ability to discover new antibiotics[[13]](https://www.thinkglobalhealth.org/article/health-latin-america-and-promise-artificial-intelligence#:~:text=Brazil%20currently%20leads%20the%20region,digital%20health%2C%20including%20AI%20applications). These examples show local talent and policy backing in the Global South, not just in Silicon Valley.  
- **Real-world example (global):** In Canada, hospitals are testing AI “early warning” systems that monitor patient vitals and alert staff to danger signs of sepsis or cardiac arrest hours in advance. In the UK’s NHS, AI tools screen retinal images for diabetic eye disease, catching cases doctors might miss. Such deployments hint at fewer medical errors and more preventative care.  
- **Challenges and caveats:** **Patient safety** is paramount – any AI error can cost lives. We must ask: How do we verify an AI’s diagnosis? Who is liable if it makes a mistake? Also, **data privacy** is critical because health data is sensitive; AI systems need vast patient data to learn, but misuse could violate privacy. **Bias** is a concern too: if an AI is trained mostly on European patient data, will it perform well for Colombian or diverse populations? There are already cases of AI misdiagnosing people with darker skin or from minority groups due to biased training data. Regulatory bodies like the U.S. FDA and EU regulators are developing approval processes for AI-based medical devices, requiring rigorous clinical trials similar to new drugs.  
- **Public trust:** Ultimately, AI won’t help if doctors and patients don’t trust it. It’s encouraging that nearly **70% of Latin American doctors** in a recent survey said AI has had a positive impact on medicine and society so far[[14]](https://globalhealthintelligence.com/ghi-analysis/how-ai-impacts-the-latin-american-healthcare-market/#:~:text=Overall%2C%20optimism%20over%20the%20impact,and%20on%20medicine%20thus%20far). However, they also feel they need more training to use it effectively[[15]](https://globalhealthintelligence.com/ghi-analysis/how-ai-impacts-the-latin-american-healthcare-market/#:~:text=Interestingly%2C%20doctors%20did%20not%20have,AI%20effectively%20in%20their%20practice). Patients, on the other hand, may be wary – a global survey showed less than half of patients might feel comfortable with an AI involved in their care. Building trust will require demonstrating success stories and maintaining a strong human touch in healthcare.  
**Recommended Visual Aids:**  
- **Statistics graphic:** e.g. a slide showing survey stats: *“55% of Latin American doctors already use some form of AI”*, *“67% believe AI has had a good impact on medicine”*[[16]](https://globalhealthintelligence.com/ghi-analysis/how-ai-impacts-the-latin-american-healthcare-market/#:~:text=%2A%2055,very%20necessary%20or%20quite%20necessary), etc. This emphasizes both the adoption and optimism among physicians.  
- **Before-and-after example:** An image or diagram of an AI analyzing a medical image (like an X-ray with AI-highlighted regions) to show how AI assists in diagnosis. Perhaps a comparison of a normal vs. AI-assisted detection of a tumor.  
- **Map or bar chart:** Indicating countries with notable health AI initiatives (highlight Brazil, Colombia, etc.) or funding in AI health. Could also show growth projections (e.g. Latin America’s AI health market growing at >30% CAGR[[17]](https://globalhealthintelligence.com/ghi-analysis/how-ai-will-impact-the-latin-american-healthcare-market/#:~:text=How%20AI%20Will%20Impact%20the,%E2%80%93%20Forecasted%20boost%20to), or a global figure like “AI in healthcare to be a $36B market by 2025”).  
**Engagement Element:** Present a short scenario: *“A rural hospital is using an AI to diagnose chest scans. One day, the AI misses a subtle sign of cancer that a specialist later catches – fortunately in time. What should the hospital do? Rely on the AI less? Retrain it? Inform all patients when AI is used?”* Ask the audience for thoughts on how to balance using advanced tools with caution. Alternatively, a direct question: “**Would you trust an AI’s diagnosis or second opinion on your condition?**” (Poll the room: e.g., *Yes, if validated / Only with a doctor’s confirmation / Not at all*). Use this to spark reflection on the importance of validation and human oversight in medical AI.

## Segment 4: **Autonomous Vehicles – Steering Toward an AI-Driven Future** (15 minutes)

*Figure: A Waymo autonomous car being tested on public roads. Self-driving vehicles promise greater safety and convenience, but they also pose new regulatory and ethical challenges.*

**Focus:** The rise of self-driving cars and AI-driven transportation. Discuss how autonomous vehicles (AVs) could improve mobility and safety, while examining the policy, safety, and societal questions they introduce.  
**Main Ideas:**  
- **What are autonomous vehicles?** Introduce the concept of AVs – cars, buses, or even trucks that can drive themselves using AI algorithms, sensors, and maps. They range from Level 2 or 3 (partial autonomy, where a human driver still needs to intervene) up to Level 5 (fully driverless in all conditions). This technology is no longer sci-fi: pilot programs are underway in several countries.  
- **Potential benefits:** **Safety is the biggest promise** – human error causes most accidents, so in theory an AI that never gets tired or distracted could save lives. AVs could also expand mobility for those who can’t drive (elderly, disabled), reduce traffic (through efficient driving and platooning), and cut emissions if coupled with electric vehicles and smart routing. For public administrators, fewer accidents could mean lower healthcare and law enforcement burdens, and smarter traffic flow could boost productivity.  
- **Global progress:** Highlight where AVs are already on the roads. In the **United States**, companies like Waymo have deployed robo-taxis – *in Phoenix, Arizona you can already hail a car that has no human driver at all*[[18]](https://robotsguide.com/robots/waymo#:~:text=In%20Phoenix%2C%20Ariz,anyone%20in%20the%20driver%27s%20seat). Some cities in California are also running autonomous ride-hailing pilots. **Europe** is testing autonomous shuttles and trucks in controlled settings; for example, Germany has legalized certain autonomous driving features on highways (with a safety driver ready). **China** is aggressively pushing AVs in cities like Shenzhen (with Baidu’s Apollo project), though under heavy regulatory oversight. **Canada** and others have authorized testing in specific zones. Importantly, Latin America is also joining in: **Chile** launched the first autonomous public transport shuttle in the region – a 3-month pilot in Santiago that carried ~2,500 passengers on a short route[[19]](https://easymile.com/success-stories/santiago-chile#:~:text=Covering%20the%20500%20meters%20from,to%20run%20in%20Latin%20America). This pilot, supported by the Chilean government and Inter-American Development Bank, shows local interest in leapfrogging into new mobility tech. Brazil has a research consortium (SegurAuto) working on AV technologies with universities and automakers[[20]](https://revistapesquisa.fapesp.br/en/self-driving-cars-the-future-of-mobility/#:~:text=One%20of%20the%20biggest%20ongoing,DAF%20Caminh%C3%B5es%2C%20and%20Vector%20Informatik).  
- **Safety and incidents:** Not everything has been smooth on the road to autonomy. There have been notable accidents involving self-driving tech – e.g., an Uber test vehicle tragically struck a pedestrian in 2018, and Teslas on “Autopilot” have collided with obstacles when the AI failed to recognize them. *Case:* Tesla’s vision-only approach struggled with detecting emergency vehicle lights, leading to **dozens of crashes with parked ambulances and police cars** in the U.S.[[21]](https://revistapesquisa.fapesp.br/en/self-driving-cars-the-future-of-mobility/#:~:text=Tesla%20cars%2C%20which%20only%20use,in%20collisions%20with%20first%20responders). These incidents underline that current AI still has blind spots. Discuss how companies are responding (e.g., adding more sensors, improving algorithms) and how regulators are reacting (some halted testing programs, now impose stricter safety checks).  
- **Ethical and legal dilemmas:** Ask the audience to consider scenarios like the “trolley problem” for AVs – if an autonomous car must *choose* between two bad outcomes (hitting one person vs. another), how should it be programmed to decide? Who bears responsibility for an AV’s actions – the owner, the manufacturer, or the AI itself? These questions are not hypothetical: legal frameworks are being updated. (Germany, for instance, has stated an AV must not discriminate (value all human lives equally) in an accident scenario – essentially forbidding it from algorithmically choosing one person over another based on age or other factors.) Liability is a major question: if a driverless car causes a crash, current laws might hold the “driver” accountable – but if there’s no human driver, it could shift liability to the company or the software. Public policy needs to evolve to assign responsibility clearly and ensure victims are compensated.  
- **Public adoption and trust:** Gauge how comfortable people are with the idea of driverless vehicles. Surveys show mixed feelings – many like the idea of safer roads, but are nervous about riding in a car with no driver. Cultural context matters too: in some countries traffic is very chaotic (imagine an AI car in the busy, rule-bending traffic of Bogotá or São Paulo). How well would today’s AVs handle that? Building trust may require demonstrating clear safety advantages. Policymakers might consider incremental introduction – e.g., allowing autonomous buses on fixed routes or trucks on highways at off-peak hours first, before general city driving.  
- **Regulation in practice:** Mention how different jurisdictions are approaching AV rules. The **EU** and UK are updating traffic laws to accommodate self-driving features (like allowing drivers to take hands off the wheel in traffic jams under certain conditions). **Brazil’s proposed AI framework explicitly lists autonomous vehicles as “high-risk AI”** that will be permitted but tightly regulated with safety oversight[[7]](https://www.mattosfilho.com.br/en/unico/framework-artificial-intelligence-senate/#:~:text=infrastructure%20,of%20affected%20individuals%20or%20groups). This means Brazil anticipates AVs and is ensuring law will require things like black box recorders or mandatory safety certifications. In the US, some states (Arizona, California) welcome tests, whereas others are cautious. **Colombia** and neighbors might look to international standards (perhaps adopting regulations once the technology is more mature).  
**Recommended Visual Aids:**  
- **Photo/Video:** A short clip or image of an autonomous car in action (e.g., a Waymo car driving in a city, or the Santiago shuttle). Seeing the vehicle with sensors and no driver can make it tangible for the audience. *(Ensure any video is brief given time.)*  
- **Infographic:** The SAE **levels of driving automation** (Level 0 to 5) diagram – helps explain the gradations from driver assist to full autonomy in a simple visual way.  
- **Statistic or chart:** Perhaps a bar chart of “Autonomous miles driven” by companies (Waymo, Tesla, etc.) to show scale of testing, or a pie chart of survey responses to “Would you ride in a self-driving car?” to illustrate public sentiment. If available, a stat like “By 2030, X% of vehicles might have Level 4 autonomy” (just be sure it’s from a credible source).  
**Engagement Element:** Ask a forward-looking question: *“If a fully self-driving bus were deployed in Manizales, how would you feel about it? Would you ride it?”* Take a quick show of hands or a couple of shout-outs. Follow with: *“What conditions would make you comfortable? For example, if it had a human safety operator? Or after it’s run accident-free for a year?”* This encourages the audience to actively consider the trade-offs and personal comfort, linking to the need for public trust and regulatory assurances. Another quick scenario: *“An autonomous car gets into an accident with a human-driven car. Who do you think should be legally liable?”* – then reveal how the law is currently unclear or evolving, which reinforces the earlier point that policy needs to catch up.

## Segment 5: **Algorithmic Fairness – Ensuring AI is Just and Inclusive** (15 minutes)

**Focus:** Address the issue of bias and fairness in AI systems. Explain how AI can inadvertently discriminate or produce unequal outcomes, and discuss strategies (technical and policy) to make AI more fair. This segment appeals to the ethical responsibility of AI in society, tying especially to public sector use (where fairness is paramount).  
**Main Ideas:**  
- **What is algorithmic bias?** Clarify that AI systems learn from data – and if that data reflects historical biases or inequalities, the AI can *carry those forward* or even amplify them. “Algorithmic fairness” means making sure AI decisions do not systematically disadvantage any group (by race, gender, socioeconomic status, etc.). Example to grab attention: facial recognition tech has shown dramatic bias – in one study, it was **nearly perfect for identifying light-skinned men but misidentified dark-skinned women 34% of the time**[[22]](https://www.aclu-mn.org/en/news/biased-technology-automated-discrimination-facial-recognition#:~:text=Studies%20show%20that%20facial%20recognition,women%2C%20children%2C%20and%20elderly%20individuals). Imagine the impact: a facial recognition system might fail to recognize an individual just because of their skin tone, or worse, falsely match someone to a crime. This has already led to wrongful arrests of Black individuals in the U.S. due to faulty AI matches.  
- **Areas of concern:** Give a quick tour of where fairness issues have surfaced:  
- **Hiring and HR:** AI resume screeners that favored male candidates because they were trained on past data where more men were hired (e.g., the Amazon hiring tool case).  
- **Criminal justice:** Predictive policing or sentencing AIs that put minority communities under heavier suspicion because historical crime data is biased or reflects discriminatory policing. (E.g., the COMPAS recidivism algorithm in the US was found to wrongly flag Black defendants as high risk more often than white defendants.)  
- **Finance:** AI-driven credit scoring or lending that might inadvertently give lower credit scores to certain ethnic groups or neighborhoods (often proxies for race or income) due to biased training data.  
- **Healthcare:** We see some algorithms that prioritize care for certain patients over others because they learned from data that underserves minorities.  
These examples underscore that **AI is not inherently neutral** – without deliberate fairness checks, it can perpetuate human biases under a veneer of objectivity.  
- **Why this matters for public administrators:** Governments are increasingly using algorithms for things like distributing social services, assessing school performance, or even deciding who gets inspected or audited. If these systems are biased, they can lead to unfair access to resources or unjust scrutiny on certain populations. It’s crucial for public sector AI to be audited for fairness – fairness isn’t just a “nice-to-have,” it’s a legal and moral requirement in governance.  
- **Tackling the problem:** Explain approaches to ensure fairness:  
- **Diverse data & testing:** Developers need to train AI on diverse datasets and then test outcomes for different demographic groups. For instance, if an AI is for loan approvals, test that its approval rates don’t unjustifiably differ by gender or ethnicity without a valid reason.  
- **Bias audits and transparency:** Some jurisdictions now require regular **bias audits** of AI systems. *Example:* New York City passed a law requiring annual bias audits of automated hiring tools and that the results be public[[9]](https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md#L100-L103). This kind of policy forces companies to check their algorithms and opens them to external scrutiny. **Canada**’s public sector AIA (mentioned earlier) also includes checking for bias as systems are evaluated[[5]](https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md#L86-L89). **Brazil**’s draft AI law gives individuals the right to not be discriminated against by AI and even to have biased decisions corrected[[23]](https://www.mattosfilho.com.br/en/unico/framework-artificial-intelligence-senate/#:~:text=,or%20indirect%20discriminatory%20bias%20corrected), embedding fairness into law.  
- **Human oversight and appeal:** Many AI governance plans (EU, etc.) mandate that when AI makes important decisions about people, humans must be able to review or override them[[8]](https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md#L134-L142). For example, if an algorithm denies someone welfare benefits, that person should have the right to an explanation and a manual review – preventing a faceless algorithm from having the final say in someone’s life.  
- **Technical fixes:** On the tech side, there’s growing research on “fair ML” – algorithms that can adjust models to reduce bias (like re-weighting data, or adversarial debiasing techniques). However, these are not foolproof, and they often involve trade-offs (improving fairness might slightly reduce accuracy, for instance).  
- **Real-world progress:** Mention that big tech companies now have ethics teams and sometimes shelve projects found to be biased. International standards bodies (like IEEE or ISO) are working on AI ethics guidelines. The EU AI Act will require high-risk AI to demonstrate lack of “unfair bias.” Also highlight local or regional awareness: for instance, the **Colombian Constitutional Court recently ruled** that algorithmic transparency is a fundamental right in public decision-making (a landmark 2025 case), signaling the judiciary’s stance that citizens should be able to scrutinize automated decisions. Latin American civil society (e.g., NGOs, academia) is active in this space – mapping AI harms in the region and pushing for inclusive AI design.  
- **Inspire responsible AI:** The takeaway is that *we* – as developers, policymakers, users – must consciously bake in fairness. AI should be a tool to **reduce** human bias (e.g., by focusing on objective factors) rather than a magnifier of it. Achieving this isn’t automatic: it requires effort, diverse teams building AI, and continuous oversight. The audience, especially future public administrators, should see ensuring algorithmic fairness as part of their mandate for technology governance.  
**Recommended Visual Aids:**  
- **Striking statistic:** Display the facial recognition error rate disparity as a simple graphic[[22]](https://www.aclu-mn.org/en/news/biased-technology-automated-discrimination-facial-recognition#:~:text=Studies%20show%20that%20facial%20recognition,women%2C%20children%2C%20and%20elderly%20individuals) (e.g., two faces icon, one labeled “light-skinned male – 0.8% error” vs “dark-skinned female – 34% error”). This drives home the existence of bias in a visual, memorable way.  
- **Headline or case photo:** A news headline about biased AI (e.g., “Algorithm Disqualifies Qualified Women Applicants” or “Wrongful Arrest Highlights Dangers of Biased Facial Recognition”). This makes the issue concrete and urgent.  
- **Policy snapshot:** A slide listing a few enacted or proposed fairness regulations (NYC bias audit law, EU AI Act fairness clause, etc.) and maybe a quote from an official like *“AI must not unlawfully discriminate”*. This shows the audience that governments are responding, and it reinforces the role of policy.  
**Engagement Element:** Prompt the audience with a mini-exercise: *“Think of an AI application you’ve heard of or used – it could be as simple as a music recommendation or as serious as a credit score. What kind of bias might sneak into it?”* Give a concrete prompt: e.g., “Does your music app suggest songs only in English? Does an AI tutor program cater less to students from rural backgrounds due to lack of data?” Have one or two people share their thoughts. This personalizes the issue of bias. Alternatively, ask a direct question: *“If you found out a hiring algorithm filtered you out because of something unrelated to your merit (say, it gave lower scores to graduates of your university), how would you react? Who should be held accountable?”* Use the responses to reinforce why transparency and avenues for appeal are necessary in algorithmic decisions.

## Segment 6: **AI and Society – A Vision for the Future + Q&A** (15 minutes)

**Focus:** Broader societal implications of AI and wrapping up with a forward-looking perspective. This segment ties together themes – emphasizing the need for **responsible public policy and informed citizenry** to shape AI’s impact. It will conclude with time for Q&A and open discussion, encouraging the audience to reflect on their role in an AI-driven world.  
**Main Ideas:**  
- **Workforce and economy:** AI is poised to transform the job market and economy. Some jobs will be *augmented* by AI, others might be *automated away*, and entirely new roles will emerge. For instance, the World Economic Forum estimated that by 2025, **85 million jobs may be displaced by automation, but 97 million new jobs could be created** in areas like data analysis, AI maintenance, and the green economy[[24]](https://www.weforum.org/press/2020/10/recession-and-automation-changes-our-future-of-work-but-there-are-jobs-coming-report-says-52c5162fce/#:~:text=,jobs%20in%20next%20five%20years). Emphasize that it’s not doom and gloom of mass unemployment *if* we manage the transition well – it’s more of a big shift requiring reskilling and education. Ask: Is the education system in Colombia and elsewhere preparing students for this AI-infused future (e.g., teaching data literacy, critical thinking, adaptability)? Governments will need forward-thinking labor policies (social safety nets, retraining programs) to ensure the benefits of AI-driven productivity are widely shared.  
- **Ethical use and public trust:** AI applications like **deepfakes** (AI-generated fake videos) and misinformation bots pose risks to society by eroding trust in information. We’ve seen examples of fake AI-generated media causing confusion in elections or spreading false news. This is a global challenge – e.g., the EU has included provisions about labeling AI-generated content to combat misinformation. On the flip side, AI can also help fact-check or detect fakes. Society will need robust media literacy and perhaps new regulations to handle this. Public trust in AI systems is fragile: polls show that while people appreciate AI’s convenience, a significant number worry about misuse – in Latin America, **44% are concerned AI could spread false information and 38% fear it will take jobs**[[25]](https://americasmi.com/insights/latin-americans-ai-marketing-trust/#:~:text=Latin%20American%20consumers%20%20see,to%20consumer%20perception%20surveys%3A%2058). Globally, about *6 in 10 consumers say trust will be the key determinant in whether they embrace AI*[[26]](https://americasmi.com/insights/latin-americans-ai-marketing-trust/#:~:text=Marketers%20should%20take%20note%2C%20because,5). Building trust requires transparency (people want to know when AI is involved and how it works), accountability (clear recourse if something goes wrong), and human values guiding AI use.  
- **Social changes and inequality:** AI could either widen or bridge social inequalities. For example, will AI-powered education tools help rural and disadvantaged communities get access to quality instruction – or will those tools be available only to affluent schools, widening the gap? There’s concern that countries or groups with more AI resources will leap ahead (“AI divide”). On the other hand, open-source AI and cheaper computing could democratize benefits. It’s partly a policy choice: governments can encourage AI for social good – like using AI to improve agriculture yields for small farmers, or to optimize traffic and reduce pollution in congested cities. We should highlight positive social uses: e.g., an AI project in Colombia that helps identify areas for reforestation or monitor Amazon rainforest illegal logging via satellite images – these show AI’s power for public good.  
- **Cultural and local context:** AI systems often carry the cultural imprint of their creators. Ensuring AI respects cultural diversity and local values is important. For instance, an AI virtual assistant might not understand Spanish idioms or might lack knowledge of Colombian cities if it’s primarily trained on English data from North America. Latin American AI researchers are aware of this and call for more local data and talent development. This is a call to action for countries like Colombia to invest in their own AI development and guide it in alignment with local needs and ethics (rather than just importing whatever big tech provides).  
- **Inspiring responsible AI policy:** Summarize that AI is a powerful tool – it can help us achieve development goals, improve quality of life, and solve pressing problems (from healthcare to climate change), *if* we handle it wisely. The audience, especially the public administrators present, have a role to craft policies that **encourage innovation** (so we don’t miss out on AI’s benefits) **while protecting citizens** (through strong ethical guidelines, education, and oversight). A phrase to leave them with might be: *“We must shape AI before it shapes us.”* The future isn’t predetermined; it depends on choices made by governments, communities, and individuals today. Colombia and Latin America can learn from global examples but also lead with their own vision (for instance, a human-rights-centric AI policy, as some have suggested for the region).  
- **Call to action:** Encourage the students to get involved – whether it’s pursuing AI ethics in their studies, participating in public discussions, or simply staying informed. And urge the public sector folks to proactively update regulations, invest in AI literacy for public servants, and ensure AI deployments are transparent and fair. The goal is an **inclusive, human-centered AI era** – where technology serves humanity, not the other way around.  
**Recommended Visual Aids:**  
- **Global vs Local snapshot:** A slide with two halves – one showing a global statistic (like the jobs numbers or a chart of AI investment globally), and another showing a local stat (like “65% of Latin Americans use AI – with Colombia leading at 59% adoption”[[27]](https://americasmi.com/insights/latin-americans-ai-marketing-trust/#:~:text=According%20to%20surveys%20by%20Bain,57)). This juxtaposition reinforces that Latin America is very much part of the AI story, both as users and innovators, while facing similar concerns about jobs and trust.  
- **Imagery of future AI society:** Perhaps an illustration of a future city with AI (drones, autonomous shuttles, people using AR glasses, etc.) to spark imagination about what embracing AI could look like – and ask whether this is the society we want and how to guide it.  
- **Policy roadmap graphic:** A simple checklist or framework for responsible AI (e.g., 1. Ethical guidelines, 2. Education & skills, 3. Regulatory sandboxes for innovation, 4. Protection against harms…). This can synthesize the lecture’s lessons into a take-home message for policy action.  
**Engagement Element:** Conclude with a reflective question to the audience: *“Having heard these examples and ideas, what* *one area* *do you feel Colombia (or your own community) should focus on to ensure AI benefits society? Is it education, new laws, investing in tech, public awareness…?”* Invite a couple of quick responses from volunteers. This gets them thinking about applying the lecture to their context.

Finally, transition to **Q&A**: open the floor for questions. Encourage anything from clarifications (“no dumb questions here, AI is a broad topic!”) to opinions (“do you agree with the EU’s strict approach or the US’s lighter approach?”) to hypotheticals (“what about AI in education or public security?” if those didn’t come up). This discussion period will allow the audience to engage directly and critically – fulfilling the lecture’s goal to *raise awareness, provoke critical thinking, and inspire responsible action* on AI.

[[1]](https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md" \l "L20-L28) [[2]](https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md#L32-L40) [[3]](https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md#L40-L43) [[4]](https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md#L76-L84) [[5]](https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md#L86-L89) [[8]](https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md#L134-L142) [[9]](https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md#L100-L103) news-ai-regulation-future.md

<https://github.com/erichecan/ctrlvai/blob/6d0a53e5dbee8683e673f5427e619e70dc7a5534/blogs/news-ai-regulation-future.md>

[[6]](https://www.mattosfilho.com.br/en/unico/framework-artificial-intelligence-senate/#:~:text=On%20December%2010%2C%202024%2C%20the,into%20law%20and%20take%20effect) [[7]](https://www.mattosfilho.com.br/en/unico/framework-artificial-intelligence-senate/#:~:text=infrastructure%20,of%20affected%20individuals%20or%20groups) [[23]](https://www.mattosfilho.com.br/en/unico/framework-artificial-intelligence-senate/#:~:text=,or%20indirect%20discriminatory%20bias%20corrected) Regulatory framework for artificial intelligence passes in Brazil's Senate - Mattos Filho

<https://www.mattosfilho.com.br/en/unico/framework-artificial-intelligence-senate/>

[[10]](https://www.thinkglobalhealth.org/article/health-latin-america-and-promise-artificial-intelligence#:~:text=,in%20Latin%20America) [[11]](https://www.thinkglobalhealth.org/article/health-latin-america-and-promise-artificial-intelligence#:~:text=But%20interest%20in%20health,patient%20engagement%2C%20and%20expanding%20telemedicine) [[12]](https://www.thinkglobalhealth.org/article/health-latin-america-and-promise-artificial-intelligence#:~:text=Awareness%20of%20AI%27s%20potential%20for,for%20the%20region%27s%20urban%20health) [[13]](https://www.thinkglobalhealth.org/article/health-latin-america-and-promise-artificial-intelligence#:~:text=Brazil%20currently%20leads%20the%20region,digital%20health%2C%20including%20AI%20applications) Health, Latin America, and the Promise of Artificial Intelligence | Think Global Health

<https://www.thinkglobalhealth.org/article/health-latin-america-and-promise-artificial-intelligence>

[[14]](https://globalhealthintelligence.com/ghi-analysis/how-ai-impacts-the-latin-american-healthcare-market/#:~:text=Overall%2C%20optimism%20over%20the%20impact,and%20on%20medicine%20thus%20far) [[15]](https://globalhealthintelligence.com/ghi-analysis/how-ai-impacts-the-latin-american-healthcare-market/#:~:text=Interestingly%2C%20doctors%20did%20not%20have,AI%20effectively%20in%20their%20practice) [[16]](https://globalhealthintelligence.com/ghi-analysis/how-ai-impacts-the-latin-american-healthcare-market/#:~:text=%2A%2055,very%20necessary%20or%20quite%20necessary) How AI Impacts the Latin American Healthcare Market | Global Health Intelligence

<https://globalhealthintelligence.com/ghi-analysis/how-ai-impacts-the-latin-american-healthcare-market/>

[[17]](https://globalhealthintelligence.com/ghi-analysis/how-ai-will-impact-the-latin-american-healthcare-market/#:~:text=How%20AI%20Will%20Impact%20the,%E2%80%93%20Forecasted%20boost%20to) How AI Will Impact the Latin American Healthcare Market

<https://globalhealthintelligence.com/ghi-analysis/how-ai-will-impact-the-latin-american-healthcare-market/>

[[18]](https://robotsguide.com/robots/waymo#:~:text=In%20Phoenix%2C%20Ariz,anyone%20in%20the%20driver%27s%20seat) Waymo - ROBOTS: Your Guide to the World of Robotics

<https://robotsguide.com/robots/waymo>

[[19]](https://easymile.com/success-stories/santiago-chile#:~:text=Covering%20the%20500%20meters%20from,to%20run%20in%20Latin%20America) Autonomous shuttle in Chile

<https://easymile.com/success-stories/santiago-chile>

[[20]](https://revistapesquisa.fapesp.br/en/self-driving-cars-the-future-of-mobility/#:~:text=One%20of%20the%20biggest%20ongoing,DAF%20Caminh%C3%B5es%2C%20and%20Vector%20Informatik) [[21]](https://revistapesquisa.fapesp.br/en/self-driving-cars-the-future-of-mobility/#:~:text=Tesla%20cars%2C%20which%20only%20use,in%20collisions%20with%20first%20responders) Self-driving cars—the future of mobility : Revista Pesquisa Fapesp

<https://revistapesquisa.fapesp.br/en/self-driving-cars-the-future-of-mobility/>

[[22]](https://www.aclu-mn.org/en/news/biased-technology-automated-discrimination-facial-recognition#:~:text=Studies%20show%20that%20facial%20recognition,women%2C%20children%2C%20and%20elderly%20individuals) Biased Technology: The Automated Discrimination of Facial Recognition | ACLU of Minnesota

<https://www.aclu-mn.org/en/news/biased-technology-automated-discrimination-facial-recognition>

[[24]](https://www.weforum.org/press/2020/10/recession-and-automation-changes-our-future-of-work-but-there-are-jobs-coming-report-says-52c5162fce/#:~:text=,jobs%20in%20next%20five%20years) Recession and Automation Changes Our Future of Work, But There are Jobs Coming, Report Says > Press releases | World Economic Forum

<https://www.weforum.org/press/2020/10/recession-and-automation-changes-our-future-of-work-but-there-are-jobs-coming-report-says-52c5162fce/>

[[25]](https://americasmi.com/insights/latin-americans-ai-marketing-trust/#:~:text=Latin%20American%20consumers%20%20see,to%20consumer%20perception%20surveys%3A%2058) [[26]](https://americasmi.com/insights/latin-americans-ai-marketing-trust/#:~:text=Marketers%20should%20take%20note%2C%20because,5) [[27]](https://americasmi.com/insights/latin-americans-ai-marketing-trust/#:~:text=According%20to%20surveys%20by%20Bain,57) How Latin Americans are using AI | Americas Market Intelligence

<https://americasmi.com/insights/latin-americans-ai-marketing-trust/>