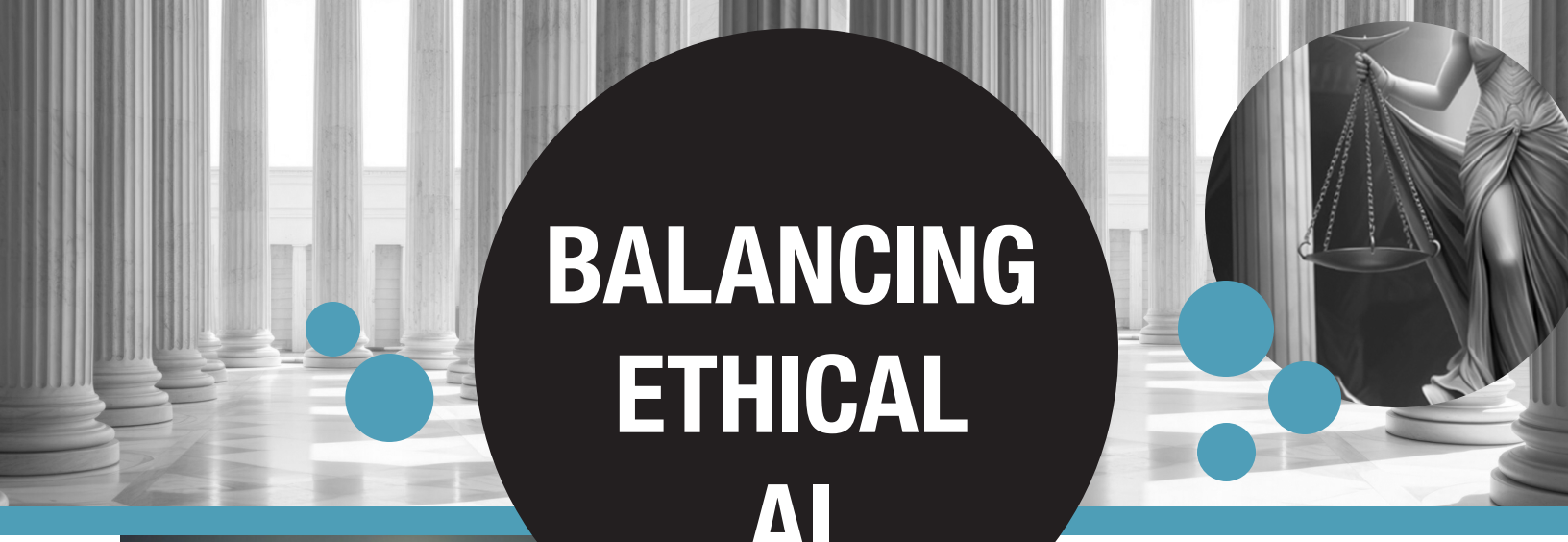

BALANCING ETHICAL AI

AI innovations and deployments are tipping the ethical balance of our society. The AI-mediated world comes with a new set of biases and unintended consequences. It's time to prioritize Ethical AI, as businesses mirror societal values and leaders strive to do the right things. I propose a novel approach that weaves the strands of AI ethics and fairness — **Five Touchstones of Ethical AI.** The framework can guide the ethical balancing act in designing AI systems as it delineates the key ethical judgements that set the standards of fairness in data, algorithms, and user feedback.

Let's calibrate and illustrate this ethical framework with a multi-faceted societal issue: the Supreme Court's recent Affirmative Action verdict. This decision, involving the University of North Carolina and Harvard College, provides an ethical backdrop to examine the robustness of the Five Touchstones. In this landmark ruling, the Justices' opinions echoed the AI principles of treating individuals equitably and considering scenarios of societal fairness. The justices' opinions, reshaped the public debate on *individual opportunity vs. societal benefits* and redefined the concept of equality highlighting the multifaceted nature of fairness, recognizing the impact of historical disparities, and the need for a nuanced approach to achieve true equality. Let's explore the ethical underpinnings of this pivotal legal decision using select AI metrics of fairness.

BASED ON INDIVIDUAL ATTRIBUTES

1. **FAIRNESS THROUGH AWARENESS:** *Compares outcomes for similar individuals irrespective of group membership, focusing on equal respect and dignity, but requiring a clear and consistent definition of similarity.* The Court's decision supports the principle of individual fairness, which requires similar treatment for individuals with similar qualifications, irrespective of group membership. The ruling indicates a preference for admissions decisions based on individual merits rather than group characteristics, aligning with the concept of individual fairness. ***Quotable quote: the touchstone of an individual's identity should be based on challenges bested, skills built, or lessons learned.***

2. **FAIRNESS THROUGH UNAWARENESS:** *This metric views an algorithm as fair, provided it does not explicitly use any protected attributes in its decision-making process.* It requires decisions to be the same regardless of an individual's race or gender. The Court's ruling that race should not be a determining factor in admissions aligns with this idea, emphasizing that an applicant's race should not change their chances of admission. ***Quotable quote: race should never be used as a negative or operate as a stereotype.***

BASED ON MEMBERSHIP IN SUB-GROUPS

3 **FAIRNESS IN RELATIONAL DOMAINS:** T*his fairness concept captures the relational structure in a domain by considering not only the attributes of individuals but also the social and organizational connections between them.* The Court recognized the importance of considering the wider context of an applicant's life, which includes their relationships and the impact of their social and familial connections on their opportunities. However, the majority opinion suggests that they must not override the principles of individual merit and nondiscrimination mandated by the Constitution. ***Quotable quote: Constitution permits but does not require them to value John's identity as a child of UNC alumni; and it permits but does not require them to value James's race—not in the abstract but as an element of who he is.***

BASED ON MEMBERSHIP IN GROUPS

4 **DISPARATE IMPACT:** *Compares outcomes for different groups in terms of social welfare, aiming to maximize overall benefit or minimize harm, while requiring a measurement of each group's preferences and value*s. The Court's emphasis on not using race as a determining factor in admissions also conflicts with this metric, which focuses on equal opportunity and justice for each group. The Court's ruling suggests that admissions should not advantage or disadvantage any group based on their racial or gender characteristics. The counter arguments are clear from the dissenting opinions. ***Quotable quotes: to satisfy "strict scrutiny" universities must be able to establish a "sufficiently measurable" link between racial discrimination and educational benefits VS. in the historical context of systemic disparities race-conscious admissions programs offer a solution towards achieving racial equality in education, in line with the equal protection mandate.***

5 **DEMOGRAPHIC PARITY:** *Compares the proportion of positive* outcomes across different groups to ensure equal representation and avoid discrimination, not accounting for individual qualifications. This statistical metric aims for equal representation among different groups, but the Court found the admissions programs' consideration of race as part of their criteria to be unconstitutional based on the majority interpretation of the Equal Protection Clause. The counter arguments is clear from the dissenting opinion. ***Quotable quotes: training future leaders and promoting diversity are commendable goals that are not "sufficiently coherent" for purposes of strict scrutiny VS. "History speaks. In some form it can be heard forever. The race-based gaps that first developed centuries ago are echoes from the past that still exist today. By all accounts, they are still stark."***

The justices' opinions, stir fresh thinking about the public good and the interplay of class and race in higher education as gateways to societal opportunities. The pivotal ruling raised profound questions about the appropriateness of using group-based metrics like demographic parity and disparate impact in business world and in academia.

'Five Touchstones of Ethical AI' offered valuable insights into the ethical judgements underpinning a multi-faceted, high-impact societal decision. Similarly, it could be a powerful Ethical AI framework for balancing social goals with economic objectives, especially when ESG and DEI objectives are considered and incorporated in AI systems. To paraphrase Drucker, *management is doing things right; leadership is doing the right things, right.*

APPLYING THE FIVE TOUCHSTONES IN BUSINESS

How the leading AI players fare when measured against the yardstick of ethical fairness? Let's explore this applying the proposed framework to identify the key considerations. In strategy settings, the pivotal step is agreeing on the most relevant touchstones for each AI system.

Google's Search Algorithm: Does it ensure *individual fairness*, or do unseen biases skew the vast landscape of information it serves to the global population?

Facebook's News Feed: In shaping the digital discourse, does Facebook's algorithm promote echo chambers, challenging the principles of *fairness through unawareness?*

LinkedIn Recruitment Tools: Are these AI tools perpetuating biases under the guise of efficiency, or are they the harbingers of a new era of *individual fairness* and merit in hiring?

Amazon's Rekognition: How does this facial recognition technology stand up to tests of disparate impact and *demographic parity,* in law enforcement applications?

IBM's Watson: In its role in healthcare, does Watson's decision-making honor *fairness in relational domains,* especially in life-altering diagnoses?

Tesla's Autonomous Systems: With decisions that could be a matter of life and death, how do these systems address *fairness in relational domains* and mitigate disparate impacts?

OpenAI's ChatGPT: At the frontiers of AI-human interaction, how do they ensure *fairness across relational domains and demographic parity.*

Microsoft's Tay: What can Tay teach us about fairness through awareness in AI, considering its rapid descent into propagating negative behaviors online?

Each of the above AI systems call for ethical application of technology, where common ground should be discovered, not demanded - honest disagreement is good for the progress of ethical AI. By applying the framework, we can scrutinize their inner workings, assess their impacts, and balance the ethical decisions. This conversation isn't just academic; it's a necessary consideration for anyone involved in the design, deployment, or governance of AI systems. Let us commit to a future where AI upholds the highest standards of fairness, respecting both individual attributes and our collective societal values.

EPILOGUE:

Good strategy is grounded in *balanced judgment*, which goes far beyond analytical reasoning and statistics. The ethical core of high-stake strategic decisions involves envisioning potential outcomes, developing evaluation criteria, assessing risks, and making decisions with empathy and compassion. Arguments and counterarguments are discerned without prejudice, and honest disagreements are encouraged. Philosophically speaking, *it is the mark of an educated mind to be able to entertain thoughts without accepting them*.

Yet, the rapid rise of AI poses a challenge to this tested and proven approach. Leaders often trust AI to enhance efficiency and decision quality, believing these systems reflect and enhance our ethics. However, this reliance risks confusing data-driven 'scenario prediction' with the nuanced art of balanced judgment. This subtle but growing influence of AI on decision processes could undermine ethical integrity of future businesses, as AI algorithms driven by biased datasets could provide illusionary interpretations of patterns. This in turn could lead to untenable predictions, and recalibration of ethical compasses without human discernment.

Therefore, it is crucial for business and technology leaders to recognize the risks and infuse human judgement based on the depth of experiences, to counterbalance recommendations based on analytical AI predictions. Such a balancing act is an imperative to sustain business stakeholders and uphold society's highest ethical ideals.

Max Michaels is a champion of Ethical AI, and an investor in companies such as Open AI and Pryon. He is an MIT alumnus who has served as an executive at IBM, AT&T, and Cisco.

He can be reached at *max@cryztal.com*