Puerto Rico #2: Beneficial AGI Conference Jan 2 to 6, 2019

The **Beneficial AGI Conference**, organized by the Future of Life Institute (FLI), took place from **January 2–6, 2019**, in San Juan, Puerto Rico.

This was FLI's third major AI conference, following the 2015 Puerto Rico and 2017 Asilomar conferences, and it focused on designing artificial general intelligence (AGI) to maximize societal benefits while addressing risks like misalignment, economic disruption, and existential threats.

Held under **Chatham House Rules** to ensure candid discussions, the event built on the 2015 open letter and 2017 Asilomar Principles, aiming to refine strategies for safe AGI development as AI systems like DeepMind's AlphaGo and OpenAI's early models gained prominence.

Who Organized It: The conference was organized by the **Future of Life Institute (FLI)**, a nonprofit founded in 2014 by **Max Tegmark** (MIT cosmologist), **Jaan Tallinn** (Skype co-founder), **Viktoriya Krakovna** (DeepMind research scientist), **Anthony Aguirre** (UCSC physicist), and **Meia Chita-Tegmark** (Tufts scholar). FLI's mission is to mitigate existential risks from advanced technologies, and this event was led by Tegmark and Tallinn, with support from FLI's team, including Ariel Conn and Lucas Perry.

Funding came from sponsors such as **Jaan Tallinn**, the **Open Philanthropy Project**, and **Elon Musk**, who had supported FLI's earlier conferences and its $7 million AI safety grant program launched in 2015. The organizing committee curated an interdisciplinary agenda, incorporating insights from FLI's 2015 and 2017 conferences and its 37 funded research teams, to address technical, ethical, and policy challenges in AGI development.

When It Occurred: The Beneficial AGI Conference was held from **January 2–6, 2019**, in San Juan, Puerto Rico, at a resort venue chosen to facilitate focused, off-the-record discussions, similar to the 2015 Puerto Rico event. The conference spanned five days, with the first two days dedicated to a technical workshop for FLI's grant recipients to share progress on AI safety research, followed by 2.5 days of broader discussions on AGI's societal implications, governance, and long-term risks. The timing reflected growing urgency in the AI field, as systems like OpenAI's GPT-2 (released 2019) and DeepMind's AlphaStar (2019) demonstrated rapid advances toward general intelligence.

Who Attended: The conference convened approximately **100 invite-only participants**, including leading AI researchers, industry leaders, economists, ethicists, and philosophers. Notable attendees included **Yoshua Bengio** (Montreal Institute for Learning Algorithms, Turing Award winner), **Stuart Russell** (UC Berkeley, AI safety advocate), **Anca Dragan** (UC Berkeley, value alignment expert), **Dario Amodei** (then OpenAI, later Anthropic

co-founder), **Gillian Hadfield** (University of Toronto, AI law expert), **Wendell Wallach** (Yale, AI ethics), and **Toby Ord** (Oxford, existential risk researcher).
Representatives from industry included researchers from **Google DeepMind**, **OpenAI**, **Google Brain**, and **Vicarious**, alongside academics from MIT, Stanford, and Cambridge. Policy experts and funders like **Jaan Tallinn** and **Sam Harris** (philosopher) also attended, ensuring a mix of technical and societal perspectives. The absence of Elon Musk, a 2015 and 2017 attendee, was notable, as he had left OpenAI's board in 2018, though his influence persisted through FLI's funding.

Principal Conclusions: The conference reinforced the need for robust AI safety research and global cooperation to ensure AGI aligns with human values, producing no new formal principles but deepening the 2017 Asilomar Principles' impact.

Key outcomes included a focus on **technical safety** (e.g., developing interpretable AI systems), **value alignment** (ensuring AGI prioritizes human goals), and **governance frameworks** to prevent misuse, such as autonomous weapons or economic monopolies.

Panels, documented on FLI's website, explored AI's economic impacts (e.g., job displacement, with 69 million AI-related jobs projected by 2027), legal accountability, and the risks of an AI arms race, echoing concerns from the 2015 open letter.

The conference emphasized interdisciplinary collaboration, with discussions led by figures like Bengio and Russell advocating for preemptive safety measures before AGI's emergence, potentially by 2030–2035. Videos and slides, including Bengio's talk on aligning AI with human intent, are publicly available on FLI's site, though some sessions remained confidential.

The 2019 Puerto Rico conference solidified FLI's role as a catalyst for AI safety, bridging industry and academia at a time when AI's rapid progress—evidenced by OpenAI's $1 billion Microsoft investment (2019) and DeepMind's AlphaStar defeating StarCraft pros—heightened urgency.

Unlike the 2017 Asilomar event, which produced the 23 Principles, this conference focused on practical next steps, such as funding safety research and influencing policy, later reflected in the 2023 Bletchley Declaration.

Its legacy includes shaping discussions at government-led summits (e.g., Bletchley Park, 2023) and reinforcing the need for global standards, though critics noted the lack of concrete enforcement mechanisms. The event's outcomes continue to inform AI governance, as seen in FLI's role in the 2024 Singapore Consensus.

Sources:

- Future of Life Institute, “Beneficial AGI 2019 Conference” (https://futureoflife.org/2019/01/10/beneficial-agi-2019/)
- Future of Life Institute, “Events Work” (2019)
- Wikipedia, “Future of Life Institute” (2023)
- Web reports on AI safety conferences (e.g., *The Verge*, 2019)