The First Informal AI Conferences in Puerto Rico

The first AI conference in Puerto Rico, titled "The Future of AI: Opportunities and Challenges," was organized by the Future of Life Institute (FLI) and held from **January 2–5, 2015**, in San Juan.

This landmark event brought together leading AI researchers, industry experts, and thought leaders in economics, law, and ethics to discuss the potential benefits and risks of artificial intelligence, particularly as it approached human-level capabilities.

Held under **Chatham House Rules** to ensure candid discussions without media presence, the conference aimed to identify research directions to maximize AI's societal benefits while mitigating existential risks, such as those later articulated by figures like Geoffrey Hinton. It marked a pivotal moment in AI safety discourse, setting the stage for subsequent efforts like the 2017 Asilomar Principles.

Attendees included prominent figures from academia and industry, such as **Elon Musk** (Tesla, SpaceX), **Stuart Russell** (UC Berkeley, co-author of *Artificial Intelligence: A Modern Approach*), **Peter Norvig** (Google, Director of Research), **Erik Brynjolfsson** (MIT, co-author of *The Second Machine Age*), **Richard Sutton** (University of Alberta, reinforcement learning pioneer), **Bart Selman** (Cornell, AAAI panel co-chair), **Tom Mitchell** (Carnegie Mellon), and **Murray Shanahan** (Imperial College). Nine researchers from **Google DeepMind**, including co-founder Demis Hassabis, attended, reflecting the company's growing influence post-2014 acquisition. Other participants came from organizations like the Machine Intelligence Research Institute (MIRI), Vicarious, and the Center for the Study of Existential Risk (CSER). Funders included **Jaan Tallinn** (Skype co-founder) and the **Open Philanthropy Project**, with Musk later donating $10 million to FLI's AI safety grants program.

The **principal conclusions** centered on the need for proactive research to ensure AI's safety and societal benefits. Attendees reached a consensus that superintelligent AI posed significant challenges, including potential existential risks, and required urgent study.

This led to the drafting of an **open letter** on AI safety, signed by attendees like Musk, Stephen Hawking, and DeepMind researchers, which called for research into AI's economic, legal, and security impacts to prevent "pitfalls" like uncontrollable systems. An accompanying research priorities document outlined key areas, such as robust AI alignment and economic disruption mitigation, heavily citing MIRI's work. The letter, made public on January 12, 2015, gained thousands of signatures, including from Yoshua Bengio, and influenced global AI safety discussions.

A notable moment was **Jaan Tallinn's presentation**, which highlighted DeepMind's AI mastering the arcade game Breakout with ruthless efficiency, illustrating both AI's potential and its risks if misaligned. This underscored the conference's focus on balancing AI's transformative power—such as eradicating disease or poverty—with safeguards against unintended consequences like job displacement or autonomous weaponization. The event's electric atmosphere, as described by MIRI's Nate Soares, reflected a shift in the AI community toward taking long-term risks seriously, catalyzing initiatives like FLI's grant program and subsequent conferences.

The Puerto Rico conference is considered a foundational event in AI safety, often compared to the 1975 Asilomar Conference on Recombinant DNA for establishing ethical research standards. Its legacy includes shaping the field of AI alignment, fostering collaboration between industry and academia, and prompting commitments like Clearpath Robotics' 2014 pledge against military AI. By addressing AI's dual potential for benefit and harm, the conference laid groundwork for ongoing debates about safe AGI development, influencing policies like the EU's AI Act and discussions at later conferences.

The first AI conference in Puerto Rico, titled "The Future of AI: Opportunities and Challenges," was organized by the **Future of Life Institute (FLI)**, a nonprofit founded by **Max Tegmark**, **Jaan Tallinn**, **Viktoriya Krakovna**, **Anthony Aguirre**, and **Meia Chita-Tegmark**, focused on mitigating existential risks from advanced technologies. Held from January 2–5, 2015, in San Juan, the conference was spearheaded by FLI's leadership, particularly Tegmark, a MIT physicist, and Tallinn, a Skype co-founder, who secured funding from Tallinn and the Open Philanthropy Project to cover costs, including travel for most attendees.

The conference was convened to address the growing potential and risks of artificial intelligence, especially as AI systems like DeepMind's (acquired by Google in 2014) approached human-level capabilities. Prompted by concerns from figures like Elon Musk and Stuart Russell about uncontrolled AI's societal and existential threats, FLI aimed to unite leading AI researchers, industry experts, and ethicists to discuss safe AI development and establish research priorities. Held under Chatham House Rules to ensure open dialogue, the event sought to balance AI's promise—such as solving disease or poverty—with safeguards against risks like autonomous weapons or economic disruption, resulting in a seminal open letter on AI safety.