Arctic regions are critical in ensuring that US sovereignty is protected. Observing, understanding, and responding to issues resulting from climate change presents unique challenges. Despite the importance of the Arctic, it is not known how current or future US military AI capabilities will perform in theater, nor how to definitively *test* them to find out. Arctic AI validation and testing is limited by the complexity of AI systems in general, which is beyond manual human verification due to their sheer size but is also compounded by the remoteness and climactic variation of changing Arctic environments. This study will consider automated test generation as a tool to algorithmically determine the trustworthiness of AI algorithms when applied to Arctic, data-sparse, and unstructured environments.