© 2020 bnh.ai CC BY-NC-SA 4.0



Phase III - Containment

	 Follow previously agreed upon "Watch Characterize breadth of attack or failure 2. 1 Identify and isolate affected sy 	e:	n" vs. "Disrupt and Disconnect" plan				
Al faile	ure						
	Determine if affected AI systems are impacting additional systems	٠	Analyze input data near probability thresholds				
	Collect system logs for affected Al systems to profile: CPU/GPU	Adver	Adversarial attacker				
	☐ Disk☐ Memory☐ Network	٠	Use intrusion detection systems (HIDS/HIPS/NIDS/NIPS) to assess unauthorized assets in any affected				
	Test affected AI system API		systems:				
	Compare affected systems' outputs		, 📮 Files				
	to documented objectives		■ Network				
	Compare affected systems'		Processes				
	behavior/use to intended use and		System calls				
	constraints		Use PCAP or other network forensic				
	Analyze Al systems' training and		devices to replay old traffic and				
	input data for restricted information		identify additional affected systems				
	Verify data lineage		Identify repetitive or anomalous				
	Analyze user comments regarding		traffic for affected systems				
	the affected AI systems		Analyze logs, queries, or scripts for				
	Assess AI system input and		training or development data				
	prediction distribution drift		systems				
	Segment affected AI systems input		Verify data lineage				
	data by performance and disparate impact		Analyze AI system production scoring code				
	Assess user-appeal and		Verify version control integrity				
	operator-override capacities						
	Test affected systems' explanations						
	against simulated data						

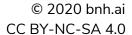
Disclaimer: bnh.ai leverages a unique blend of legal and technical expertise to protect and advance clients' data, analytics, and Al investments. Not all firm personnel, including named partners, are authorized to practice law.



Phase III - Containment (Cont.)

	3. Det	ermine l	osses							
	٥	analys		data			-		-	oint traffic, and vity of the loss and
		3.1 Ty	pe and Sensitivit	y of	f Loss:					
0	Public Image Identit Interna Metere	al docur documos ies al messa ed comp	ents ages oute		Model outcomes (e.g., loans, insurance policie promotions, etc.) Operational data Sound and video Source code	es) a	5,		<u> </u>	Statistical or ML models (encoded data and proprietary logic) Training data Other
		3.2 lm	pact of Incident:							
			Confidentiality Integrity Availability							
	4. Initial assessment of compliance and legal liabilities:									
	Fairne	ss:					Priva	су		
		Repres Insuffic	discrimination sentativeness of cient testing cient monitoring	dat	a		0	Legal	nak bas	
	Securi	ty:					Othe	:		
		Safety	ecurity security standards n reporting					Decep Warra	otiv anti ousl	y generated

Disclaimer: bnh.ai leverages a unique blend of legal and technical expertise to protect and advance clients' data, analytics, and Al investments. Not all firm personnel, including named partners, are authorized to practice law.





Phase III - Containment (Cont.)

	5. If necessary or appropriate, alert FBI or other law enforcement
	6. If necessary or appropriate, alert CFPB, FDA, FRB, FTC, or other federal regulatory body
	7. If necessary or appropriate, alert state regulators and attorneys general
	 8. If necessary or appropriate, inform public: Customer notifications Industry requirements Partner and third-party notifications US-CERT Al incident databases
.	 9. Consider and prepare for reputational harm associated with failure or attack: Internal communications Public relations & external communications Legal privilege

Disclaimer: bnh.ai leverages a unique blend of legal and technical expertise to protect and advance clients' data, analytics, and Al investments. Not all firm personnel, including named partners, are authorized to practice law.