

Non c'è crisi tecnologica nella Silicon Valley

Grandi e piccole imprese digitali pagano il prezzo della crisi finanziaria. Ma il ciclo tecnologico è pronto a ripartire con le prime applicazioni su larga scala delle tecnologie di intelligenza artificiale. Sarà la regolamentazione, figlia di una politica che inizia a occuparsi del digitale, a far capire che la crescita non è infinita. Ne parliamo con l'imprenditore e informatico Stefano Quintarelli

Giulio De Petra, Stefano Quintarelli

Si parla molto di crisi del settore digitale facendo riferimento a quantità di licenziamenti mai visti nella Silicon Valley, circa 65.000 nei primi mesi del 2023 sommando Amazon, Meta e Google. È un assestamento fisiologico dopo l'impennata di attività dovuta alla pandemia, oppure è la fine di un lungo ciclo di crescita? Siamo al punto massimo di espansione dei modelli di business dei Big Tech?

Non c'è una risposta unica che dia conto di tutte le diverse sfaccettature di quello che sta accadendo. Ma se questa è una crisi, c'è da augurarsela una crisi così. Possiamo parlare di crisi quando sei in recessione, quando perdi quattrini, e questo non sta accadendo dal punto di vista industriale se guardi al fatturato.

Per capire dobbiamo considerare l'aspetto finanziario.

Alcune aziende erano sopravvalutate perché scontavano una aspettativa di crescita infinita dei ricavi che si è sommata a una grande quantità di liquidità sui mercati che è finita in queste aziende. Flussi finanziari che si giustificavano con dei modelli previsionali azzardati.

Nell'ultimo anno le grandi quantità di risparmio accumulate durante gli anni del Covid si sono riversate sull'economia portando a un aumento di prezzi

che si è accumulato con l'aumento dei costi dell'energia provocato dalla guerra. Questo ha generato inflazione che le banche centrali hanno provato a contenere con l'aumento dei tassi di interesse. Da un mondo con tassi di interesse nulli o negativi si è passati a un mondo con tassi di interesse un po' alti. Molti investitori sono usciti dai mercati e le quotazioni si sono un po' abbassate. Ma rendersi conto che le cose non crescono all'infinito è una conseguenza sana che non chiamerei crisi.

Bisogna inoltre considerare il comportamento terribile di alcuni investitori finanziari. Ho visto direttamente la lettera di un grande fondo pensionistico che scriveva a una importante azienda della Silicon Valley in cui c'era scritto: "Questi sono i tuoi ricavi per persona, questi sono i tuoi margini per persona, noi crediamo che ci possa essere un miglioramento e quindi devi tagliare di una certa percentuale il tuo personale e siamo sicuri che farai la cosa giusta per valorizzare al meglio il tuo titolo".

La pressione sul taglio dell'occupazione non è giustificata dalla situazione industriale delle Big Tech che continuano a fare utili e dividendi, ma deriva dalla pressione dei grandi investitori finanziari. In un mondo non così finanziarizzato non ci sarebbero stati licenziamenti così significativi.

Quindi è solo ciò che accade nei mercati finanziari che spiega i licenziamenti?

Se non ci limitiamo alle aziende più grandi, ai Big Tech, ma consideriamo tutto il sottobosco di piccole e medie aziende alimentato dai fondi di *venture capital*, fino all'anno scorso molte di queste aziende avevano come missione del loro business quella di vendere sé stesse al prossimo finanziatore, piuttosto che vendere prodotti e servizi al mercato. Questo modello si è inceppato a causa dell'aumento dei tassi di interesse. È una conseguenza sana che questo tipo di aziende siano costrette a concentrarsi sugli aspetti produttivi e non su quelli finanziari, ma inevitabilmente questo significa che alcune di queste aziende, anche grandi, del sottobosco delle Big Tech siano costrette a guardare con più attenzione al loro conto economico e quindi anche a licenziare.

Per le Big Tech c'è un problema di pressione degli investitori finanziari. Le società che gli stanno intorno erano invece legate a ipotesi di business drogate dai tassi di interessi negativi e sono state costrette a fare i conti con l'economia reale.

Questo non c'entra con il ciclo tecnologico e avviene, ad esempio, in un momento di grande espansione di tutte le applicazioni che fanno uso del *machine learning*, di questa ondata tecnologica che è ancora in espansione.

La mia lettura quindi è che si tratta di un perturbazione esogena dovuta ad aspetti finanziari.

In questo quadro rientra anche il fallimento di Silicon Valley Bank?

Silicon Valley Bank è fallita perché negli USA non hanno gli stress test che abbiamo in Europa, che misurano le condizioni di resilienza delle banche. SVB era una banca fragile non tanto come patrimonio, ma come struttura di esposizione al rischio. Una banca normale ha tanti clienti con piccoli importi, e alcuni di questi clienti che chiedono prestiti. La banca prende i soldi dai correntisti, con una parte concede dei prestiti ai suoi clienti, e il resto lo investe in buoni del tesoro, azioni etc..

SVB era la banca di aziende fortemente liquide, tutto il mondo delle start up che ricevono 10/20 milioni di dollari dal *venture capital* e che li depositano in banca. I clienti di SVB erano fatti così, andavano in banca non per finanziare la loro azienda, ma per depositare i soldi dell'azienda. Se chiedevano un prestito non era per l'azienda ma per comprarsi casa. SVB a differenza della media americana aveva pochi correntisti con tanti soldi e tutti dello stesso tipo. E per guadagnare investiva in buoni del tesoro a dieci anni, con tassi di interesse bassi. Con l'aumento dei tassi di interesse le start up, come abbiamo detto, non sono riuscite a trovare nuovi e più ricchi investitori e quindi hanno dovuto ritirare i soldi per pagare la produzione. La banca per restituire i soldi ha dovuto vendere i buoni del tesoro, che però si erano deprezzati e hanno generato perdite sul conto economico. Gli indicatori della banca sono diventati negativi ed essendo una banca che funziona tutta online i depositi sono scomparsi in un attimo.

Quindi il problema di SVB non aveva a che fare con il ciclo tecnologico, ma con la particolare natura dei suoi clienti. Anche in questo caso la causa non deriva dalla tecnologia ma dall'inflazione.

Si tratta quindi di perturbazioni che nascono solo dai mercati finanziari?

C'è un altro aspetto che bisogna considerare e che riguarda la regolazione dell'antitrust che sta iniziando ad avere i suoi effetti. L'illusione della crescita infinita è terminata anche perché stanno arrivando le cause antitrust, il *digital market act*, il *digital service act*, cioè perché sta arrivando la regolamentazione che non produce degli effetti subito, ma che la finanza sconta in anticipo, Ma anche questa non è una questione tecnologica.

Ma in questo senso si può dire che c'è una relazione tra questa crisi, che tu valuti apparente, delle Big Tech e la crescita della consapevolezza sociale sui danni dei loro modelli estrattivi?

Sì, nel senso che questa crescita di consapevolezza è incorporata dalla politica nelle nuove proposte di regolamentazione.

Ma non c'è un problema di effettiva capacità di esigere il rispetto di questa regolamentazione?

Sicuramente produciamo più regolamentazione di quella che le amministrazioni riescono ad applicare. Ma questi tentativi di regolamentazione ci dicono che la politica ha scoperto la tecnologia. Se guardiamo a una produzione tradizionale, come quella delle automobili, le direttive e i regolamenti da applicare sono ben 125. Ma questo non ha impedito all'industria automobilistica di svilupparsi. E lo stesso accadrà per l'industria digitale.

Come ha operato invece la segmentazione dei mercati dovuta alle conseguenze geopolitiche della guerra? Lo sviluppo delle Big Tech ha nuotato nell'acqua della globalizzazione. Come è stato influenzato dalla fine dei mercati globali?

Per rispondere bisogna considerare caso per caso le diverse tipologie di azienda. Per esempio Tesla è l'unica grande azienda tecnologica in Cina con il 100% di proprietà straniera. È chiaro che Tesla risentirà dell'inasprirsi delle tensioni tra Stati Uniti e Cina. Google e Microsoft non ne risentono

con la stessa intensità perché è da molto tempo che sono in Cina e hanno già adeguato il loro modo di operare. C'è un effetto quindi, che contribuisce a rafforzare l'idea che la crescita non è infinita, ma non è tale da determinare una crisi, anzi è prevedibile che nella prossima trimestrale gli utili delle Big Tech aumenteranno.

Qualcuno certo perderà qualcosa, ma più per la concorrenza tra i diversi attori. Ad esempio Facebook è in una situazione complicata perché Apple, che vuole entrare nel mercato pubblicitario e ha già accordi con Google in questo senso, sta facendo la guerra a Facebook rendendogli impossibile la pubblicità 'targettizzata' e presentando questo come protezione del consumatore. Anche Amazon sta entrando nel mercato pubblicitario erodendo lo spazio di Facebook.

Hai detto che siamo alla vigilia di un nuovo ciclo tecnologico legato allo sviluppo del *machine learning* e, più in generale, alle applicazioni della intelligenza artificiale. È una discontinuità rispetto allo sviluppo che abbiamo finora conosciuto, oppure sono nuove opportunità in continuità con il ciclo precedente?

Fatico a vedere un nuovo business basato sui *Large Language Models* (LLM) che sia sostenibile e multimiliardario. Dal mio punto di vista non c'è una vera discontinuità. Non c'è la messa in crisi del mondo precedente e l'affermazione di un nuovo player. Per la natura stessa degli LLM.

Provo a dirlo con un esempio: c'è uno che fa fotocopie di una planimetria di una casa con una macchina Xerox, e la fotocopia è identica all'originale salvo che per un piccolo dettaglio, che le stanze hanno tutte la stessa metratura. Perché accade questo? Perché non si tratta di una copia fotostatica, è una acquisizione digitale, con salvataggio del file e stampa. Il problema è che il salvataggio del file ha un rapporto di compressione per cui, visto che i numeri dei metri quadri si assomigliavano tutti, la macchina ne ha compresso solo uno, e quando lo ha stampato le stanze avevano tutte la stessa metratura. Quando salviamo una immagine digitale, noi decidiamo il livello di compressione e quando successivamente ingrandiamo l'immagine ci accorgiamo, ad esempio, che i bordi non sono netti. I sistemi come ChatGPT sono un po' così, prendono una immagine linguistica, ne tirano fuori un modello, che è come una "compressione" del testo e quindi

perde dettagli; quando zoomano al suo interno si scopre che ci sono degli artefatti, che le cose non sono giuste. Se io chiedo a ChatGPT di produrre la favola di una principessa e di un metalmeccanico che sconfiggono un drago sottomarino, mi scrive una bellissima storia. Se io gli chiedo chi è Stefano Quintarelli – e glielo chiedo con o senza un accento o un punto di domanda – ottengo risposte assolutamente bizzarre perché sto andando a zoomare su un dettaglio.

Queste conseguenze sono insite in questo tipo di tecnologie, e ne limitano moltissimo gli ambiti applicativi. È uno strumento potente di espansione delle nostre facoltà nella produzione di testi, ma non ci evita di dover mettere comunque mano ai testi.

Tutto il tema del *machine learning* è in realtà un modo diverso di fare software, che è con noi e ci resterà. Abbiamo fatto il software algoritmico, abbiamo fatto le reti, adesso facciamo software “statistico”, ed è una ondata tecnologica che resterà, ma senza una grande discontinuità con il precedente mondo del business. Diversa è la questione se guardiamo agli effetti sociali.

Vedi effetti sociali diversi da quelli che abbiamo imparato a conoscere e anche a denunciare negli ultimi anni?

No, sono gli stessi. Ma il problema è che sono ancora troppo pochi quelli che ne hanno consapevolezza. Adesso le persone cominciano a considerare che non ti puoi fidare di qualunque cosa leggi su Google o su Twitter. Hanno imparato a non fidarsi del tutto. Ma con sistemi come ChatGPT sarà sempre più difficile, perché le storie prodotte sono molto più credibili, perché sono sintatticamente fatte molto bene e vengono prodotte in massa sulla base del profilo della persona. Il livello di sofisticazione della “non realtà” è molto alto e passerà molto tempo prima che le persone capiscano che non si possono fidare.

La conseguenza sarà che nel medio periodo diventerà più importante il mezzo, che diventa lo strumento di certificazione. Non potendo distinguere tra vero e non vero, tra irreale e reale, dovrò avere fiducia in chi me lo dice.

Non è qualcosa di simile a quello che è avvenuto con l'elaborazione delle immagini?

Filmati finti e fotografie finte sono molto diffusi, e molti si fanno abbindolare. Ma avere la capacità di produrre linguaggio aumenta enormemente la possibilità di generare cattiva informazione. Molto più delle immagini. Perché la quantità di materiale prodotto può essere enormemente maggiore. Inoltre produrre un testo che dice una balla, ma che sembra una cosa vera, e scriverlo in modo diverso per ogni persona, oggi è possibile. Farlo con una immagine richiede molto più lavoro e controllo umano. Sono ordini di grandezza diversi.

Quali effetti ci saranno sul mondo italiano della ricerca e delle aziende che lavorano nel settore digitale?

I Large Language Models (i sistemi come ChatGPT, ndr.) sono uno strumento poderoso che richiedono una quantità di calcolo mostruosa che non è alla portata di una azienda italiana. Per cui ci dovrebbe essere, a livello europeo, un CERN per l'intelligenza artificiale che renda disponibile alle aziende quantità di dati e capacità di calcolo adeguati e utilizzare queste tecnologie.

Le aziende italiane sono piccole perché hanno pochi dati. Ma in Italia non dipende solo da questo. Da qualche giorno è uscito sul sito del Ministero dell'Istruzione un avviso che tranquillizza le scuole che utilizzano tecnologie USA rispetto al tema della privacy, riportando come giustificazioni le affermazioni di marketing delle aziende stesse. Se confronti questo con il Ministero della Cultura francese che decide di investire in software open source, con i Ministeri e i Lander tedeschi che vietano la possibilità per le amministrazioni pubbliche di utilizzare determinati strumenti per favorire le imprese tedesche, ti rendi conto che in Italia non siamo capaci di approfondire e di trarre le conseguenze di questo approfondimento. Riusciamo a individuare un problema ma non proviamo nemmeno a risolverlo. Questa sorta di "inevitabilismo" si riverbera sulle aziende digitali italiane riducendo gli spazi di opportunità.

Auspichi un CERN della IA che renda disponibile alle aziende quantità di dati e capacità di calcolo adeguati. Ma anche le applicazioni della IA avranno conseguenze sociali?

Ti rispondo con un esempio. Se tu sei un medico e usi un sistema di intelligenza artificiale per fare una diagnosi e sei in disaccordo con quello che ti suggerisce il sistema hai comunque un incentivo a uniformarti alla diagnosi del sistema, perché se poi il tuo paziente muore puoi dire che anche il sistema di IA aveva confermato la tua analisi. Viceversa, se tu, essendo in disaccordo col sistema, privilegi la tua opinione e non confermi la diagnosi della IA, e poi il paziente muore per la tua diagnosi sbagliata, ti imputeranno di aver perseverato nella tua idea anche se il sistema aveva formulato una diagnosi diversa. Questo è un incentivo a diminuire la centralità umana nella decisione. E in alcuni ambiti questo può essere pericoloso. Come si fa a diminuire questa conseguenza sociale negativa nell'uso della IA? Bisogna progettare il sistema in modo tale che, quando formula una predizione, ogni tanto, casualmente, la predizione viene cambiata di segno.

In questo modo, nel nostro esempio, il medico sa che non si può più fidare del sistema di IA. La macchina può fare il 90% del lavoro, ma poi sei tu che devi prendere una decisione, perché è sicuro che la macchina potrebbe anche dirti il falso.